# Demography and the evolution of genetic and cultural variation

**Adam Powell**

**September 2010**

Thesis submitted for the degree of

Doctor of Philosophy

to

Research Department of Genetics, Evolution and Environment

University College London (UCL)

I, Adam Powell, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

# Abstract

This thesis addresses how demographic processes affect both genetic and cultural variation. Drawing on theory and techniques from both gene-culture coevolution and population genetics, and using both genetic and archaeological data, I present a number of projects covering a wide range of questions on the evolutionary history of our species.

Chapter 2 develops a simulation model of the gene-culture coevolution of lactase persistence and dairying in Neolithic Europe. Using approximate Bayesian computation (ABC) to integrate modern genetic and archaeological data, the results demonstrate that this coevolutionary process began ~7,500 years ago in central Europe. The inferred origin is closely associated with the emergence of the Linearbandkeramik (LBK), an early cattle-based dairying culture. Chapter 3 extends a previous cultural evolutionary model to show that heterogeneity in subpopulation density causes spatial structuring of culturally inherited skill accumulation. Genetic estimates of regional effective population density demonstrate that the appearance of modern human behaviour during the Late Pleistocene can be explained by demographic factors. Chapter 4 reviews the application of the neutral model in archaeological and other cultural contexts, and develops a novel statistical approach to test for deviation from neutrality. Results show that surprisingly high levels of non-neutrality, in the form of frequency-dependent copying, can go undetected. Chapter 5 develops a novel neutral model of cultural evolution, relaxing some previous common assumptions outlined in Chapter 4. The model allows accurate estimation of population parameters from minimal archaeological data, and is applied to a decorated pottery dataset from

the LBK settlements of southwest Germany. Chapter 6 outlines the utility of the coalescent model of population genetics in inferring potentially complex demographic histories using both modern and ancient DNA samples. I provide two examples of my work from collaborative projects on the domestication of Near Eastern cattle and the demographic history of the Hispaniolan hutia.

# Table of contents

page

# List of tables

# List of figures

# List of abbreviations

| | |
|---|---|
| ABC | approximate Bayesian computation |
| aDNA | ancient DNA |
| AMH | anatomically modern human |
| bp | base pairs |
| BSP | Bayesian skyline plot |
| CI | Bayesian credible interval |
| $CI_f$ | frequentist confidence interval |
| GB | 'genetic background' |
| GLM | general linear model |
| HGDP | Human Genome Diversity Project |
| HLA | human leukocyte antigen |
| HPD | highest posterior density |
| IS | importance sampling |
| kya | thousand years ago |
| LBK | Linearbandkeramik |
| LCT | lactase gene |
| LGM | last glacial maximum |
| LP | lactase persistence |
| LSA | Late Stone Age |
| MCMC | Markov chain Monte Carlo |
| MLE | maximum likelihood estimate |
| MRCA | most recent common ancestor |
| MSA | Middle Stone Age |
| MSE | mean square error |
| mtDNA | mitochondrial DNA |
| mya | million years ago |

| | |
|---|---|
| OIS | oxygen isotope stage |
| PC | principal component |
| PCA | principal component analysis |
| PCR | polymerase chain reaction |
| PLS | partial least squares |
| NCPA | nested clade phylogeographic analysis |
| RAO | Recent African Origin |
| rMSE | relative mean square error |
| SNP | single nucleotide polymorphism |
| $T_{MRCA}$ | time to most recent common ancestor |
| UP | Upper Palaeolithic |
| ya | years ago |

# List of software / programming languages

| | |
|---|---|
| Arlequin | Excoffier *et al.* (2005) |
| BayeSSC | Anderson *et al.* (2005) |
| BEAST | Drummond and Rambaut (2007) |
| Serial SimCoal | Anderson *et al.* (2005) |
| KernSmooth | R package (Wand and Ripley 2009) |
| numpy | Python library (Ascher *et al.*), URL:http://numpy.scipy.org/ |
| Python | van Rossum *et al.*, URL: http://www.python.org/ |
| R | R Development Core Team, URL: http://www.R-project.org |

# Acknowledgements

Thanks to all UCL friends and colleagues in and around the Centre for the Evolution of Cultural Diversity (CECD) and the Centre for Genetic Anthropology (TCGA); and in particular Krishna Veeramah, Yuval Itan, Anne Kandler, Shakti Lamba, Chris Plaster, Lauren Johnson and Sarah Browning. Thanks also to Chris Stringer, Rob Boyd, Pete Richerson and Joe Henrich for useful comments on the work contained in Chapter 3; and further thanks go to past, present and hopefully future collaborators: Alex Bentley, Peter Schauer, Joachim Burger, Pascale Gerbault, Ian Barnes, Selina Brace and Ruth Bollongino.

I owe a real debt of gratitude to both Heidi Colleran and Fiona Jordan, who have read and commented insightfully on much more of this thesis than I had any right to expect. A huge thanks also to my family – especially my parents Keith and Hilary – and friends, whose constant support and encouragement made this work possible.

My greatest thanks are reserved for my supervisors Mark Thomas and Stephen Shennan, who have provided invaluable guidance over the past 4 years. Their encouragement to explore new ideas, coupled with their complementary expertise in genetics and archaeology, has afforded me the much-appreciated freedom to work on the wide variety of topics presented in this thesis.

*This work was funded by the UK Arts and Humanities Research Council.*

# Chapter 1

# Introduction

## 1.1 Human origins

It is now widely accepted that the evidence from both genetics and palaeoanthropology (Stringer and Andrews 1988, Harpending and Rogers 2000, Stringer 2002, Ingman *et al.* 2000, White *et al.* 2003, McDougall *et al.* 2005, Stringer 2007, Fagundes *et al.* 2007) overwhelming supports the view that anatomically modern *Homo sapiens* evolved in Africa around 200 thousand years ago (kya). Modern humans expanded from the continent to colonize the Old World some time between 90 and 40 kya (Lahr and Foley 1998, Ambrose 1998a, Ingman *et al.* 2000, Ray *et al.* 2005), largely replacing *Homo erectus* in Asia and eventually the Neandertals in Europe (though see below). Debate still surrounds the number of expansion events and the routes taken (see for example Stringer 2000) but it appears the spread eastwards across Eurasia was relatively rapid, reaching what is now Australia between 40 and 60 kya (O'Connell and Allen 2004, Hudjashov 2007), and finally the Americas some time after the last glacial maximum (LGM) ~22kya (Goebel *et al.* 2008, Ray *et al.* 2010).

The expansion from the African continent has long been associated with a significant increase in technological, cultural and social complexity, deemed characteristic of modern behaviour – the Upper Palaeolithic or Late Stone Age (Bar-Yosef 2002, Mellars 2005). However, in the light of recent evidence from

Middle Stone Age sites across sub-Saharan Africa (e.g. McBrearty and Brooks 2000, Henshilwood *et al.* 2002, d'Errico *et al.* 2003) it is now less clear whether the emergence of such modern behaviour long predated this expansion (McBrearty and Brooks 2000), was an immediate trigger caused by some kind of neural genetic mutation (e.g. Klein 1999, 2000), or developed in response to the new ecological conditions encountered outside sub-Saharan Africa (Stiner and Kuhn 2006, Mellars 2005). Another line of argument suggests that demographic factors underlie the appearance of modern behaviour (Shennan 2001, Stringer 2007, Powell *et al.* 2009), and this is the focus of Chapter 3 of this thesis.

While the details remain to be settled, this broadly supported 'Recent African Origin' (RAO) model (Howells 1976, Stringer and Andrews 1988, Harpending *et al.* 1993, Lahr and Foley 1998, Stringer 2007) stands in contrast to the multiregional model, which proposes a gradual simultaneous evolution of modern humans from archaic *Homo* species across the Old World (for example Thorne and Wolpoff 1981, Wolpoff *et al.* 2000). The earliest versions of this model denied a unique African origin for modern humans, arguing that 'regional continuity' with ongoing intercontinental gene flow better explained the apparent regional differences in morphology. The recent publication of the draft Neandertal genome (Green *et al.* 2010), recreated from ancient DNA extracted from bone fragments of three individuals, has indicated that there *was* some level of interspecies gene flow following the expansion of *Homo sapiens* from Africa. These results suggest a more complicated history for our species than a model of a strict replacement of archaic *Homo* species would allow (see also Chapter 3; section 3.7.1). However, it is also overwhelmingly evident that Africa plays a central role in our evolutionary history and that an RAO model incorporating a (still very limited) degree of modern-archaic gene flow will likely provide the most complete

description of modern human prehistory (see Plagnol and Wall 2006, and also Fagundes *et al.* 2007).

Following the colonization of the Americas <~16.5 kya[1] (Goebel *et al.* 2008, Ray *et al.* 2010, Greenberg *et al.* 1986), *Homo sapiens* had spread into almost all habitable regions on Earth – except the remote islands of the Pacific – spanning a vast range of ecological niches. Living in small bands, consisting largely of families and closely related kin, we subsisted primarily as nomadic hunter-gatherers, hunting and trapping all kinds of game – from tortoises to mammoths – and exploiting a wide variety of marine and plant resources.

The stabilization of the global climate around 11.5 kya marked the end of the climatically volatile Pleistocene era – in which we spent the majority of our evolutionary history – and the beginning of the warmer Holocene era. The domestication of a range of plants and animals began during the Neolithic at a number of independent regional centres around the world, including the Fertile Crescent in the Near East ~11 kya (see Chapter 2), the Yangzi and Yellow river basins in modern-day China ~9 kya, and Mesoamerica ~5 kya. The development of agriculture and the intensification of animal domestication led to a more sedentary mode of subsistence, and the higher population densities that it could support led to the rapid growth of farming populations, which soon came to represent the overwhelming majority of the world's population (Richerson *et al*. 2001). While there is some evidence of permanent settlements prior to the onset of farming, there is nothing comparable to the first villages – then towns and cities – associated with early farming cultures. Vast changes in social organization followed, as a consequence of living in larger and more permanent communities,

[1] All archaeological dates are given as calibrated $^{14}C$ dates.

including increased division of labour and the emergence of specialization and the development of more complex hierarchies and social and religious institutions. The subsequent development of writing and the emergence of recognizable political entities such as city-states, states and empires, ushers us firmly from human prehistory into the realm of history (see for example Diamond 1998, Cavalli-Sforza *et al.* 1994, Scarre 2005).

Notwithstanding the necessary oversimplifications made in this brief sketch of human origins – that no doubt does violence to a whole range of disciplines – it is worth labouring the point that we are very clearly a species unlike any other. What was it that made *Homo sapiens* so staggeringly successful? What allowed an east African savannah ape to colonize some of the most extreme ecological niches on Earth and everything in between? A number of genetic adaptations to the novel ecological conditions we have encountered throughout our evolutionary history have been identified (e.g. skin pigmentation, hair morphology, see Voight *et al.* 2006, Williamson *et al.* 2007), but of equal – if not more – importance is the fact that we are a highly *cultural s*pecies – on a scale not even approached by any other animal. Our evolved capacity for accumulating a vast range of cultural and technological adaptations, or 'cumulative culture' (Boyd and Richerson 1985, Tomasello 1999, Richerson and Boyd 2005), has helped us to colonize and exploit every possible niche.

This seemingly infinite store of accumulated technological, ecological and cultural knowledge, gained largely through socio-cultural learning processes, is key to understanding both our spectacular success as a species and the global variation in cultural behaviours and practices. Our 'cultural toolkits' allow us to successfully negotiate living in ultra-dense modern metropolises to sparsely populated desert and tundra regions, where reservoirs of context-specific

accumulated ecological information may be vital for survival. We live within a wide continuum of political and socio-economic systems – from totalitarian military or clerical dictatorships, through liberal welfare states and hereditary chiefdoms, to the egalitarianism of some hunter-gatherer groups. We each speak one of the estimated ~7000 extant languages, display a variety of marriage and wealth inheritance practices, and worship a range of gods and deities, from none at all to vast pantheons.

While a number of disciplines have, over the past decades and centuries, provided a wealth of invaluable insight into the full spectrum of human diversity, perhaps one of the most promising is the relatively young field of gene-culture coevolution. This discipline aims to provide integrated explanations for observed human genetic and cultural variation, by seeing it as a result of the interaction between the dual inheritance systems of genetics and culture (Boyd and Richerson 1985, Cavalli-Sforza and Feldman 1981, Durham 1991, 1992, Laland *et al.* 2010, Richerson *et al.* 2010). This thesis addresses a wide range of evolutionary questions, using gene-culture coevolutionary (Chapter 2), culture evolutionary (Chapters 3, 4 and 5) and population genetic (Chapter 6) theory. While each chapter is essentially a self-contained project, the common thread running throughout is that of the critical importance of considering the effects of demography on the evolution of genetic and cultural variation. The remainder of this Introduction provides some of the necessary background and theory, before giving an overview of the rest of the thesis (section 1.4).

## 1.2 Gene-culture coevolution and cultural evolution

### 1.2.1 Evolution and human behaviour

The publication of *Sociobiology* by E.O. Wilson in 1975, which sought to extend the principles of evolutionary biology and behavioural ecology into the human domain, sparked one of the biggest and most acrimonious academic disputes in recent decades. Attacked by some evolutionary scientists, anthropologists and sociologists alike (see Allen *et al.* 1975, Rose *et al.* 1984), Wilson was accused of racism, reductionism, biological determinism, political naivety and – at the more constructive end of the spectrum – ignoring the importance of culture in influencing human behaviour (see Laland and Brown 2002). Many of these critics feared that *sociobiology* represented an attempt to resurrect some of the ideas of early 'Darwinist' anthropologists (for example Morgan 1877, Tylor 1871) – developed shortly after Darwin's original work (1859, 1871) – some of which sought to use his ideas to provide 'scientific' support for prejudice. While the widespread, and mostly undeserved, hostility towards human sociobiology effectively ended it as a viable line of research, it laid important groundwork for a number of other disciplines, with many of its ideas still thriving under a different name: evolutionary psychology.

The aims of *evolutionary psychology* are to investigate the psychological mechanisms that underpin observed human behaviour and to seek to uncover the past selection pressures that may have shaped them (see for example Barkow, Cosmides and Tooby 1992, see also Laland and Brown 2002). Evolutionary psychologists often stress the fact that environmental changes may have outpaced the rate of genetic selection, leaving us with a 'stone-age mind' in a modern world, and hence that we should expect to now observe some non-adaptive human

behaviours. However, many in this discipline have continued to ignore or downplay the diversity in and importance of transmitted culture in influencing human behaviour.

Another discipline to emerge from the study of animal behaviour is that of *human behavioral ecology*, which is concerned with how individuals flexibly modify their behaviour in response to ecological conditions to maximize their reproductive success (Smith and Winterhalder 1992, Smith 2000, see also Laland and Brown 2002). Human behavioural ecologists allow a greater role for culture, assuming that it forms part of the adaptive response to environmental changes. There is an increasing degree of cross-over between human behavioural ecology and gene-culture coevolution, but the view that culture is shaped largely by exogenous selective pressures rather than through the various inter-individual transmission processes of gene-culture coevolution, means that human behavioural ecology may struggle to explain clear examples of sub-optimal behaviour.

A further evolutionary approach to human behaviour is memetics, which draws a direct analogy with genetic evolution. The idea of a discrete cultural unit, or 'meme', that is replicated in the brains of individuals through imitiation or transmission was first introduced informally by Dawkins (1976), and further developed by a number of other authors (Dennett 1995, Blackmore 1999, Aunger 2002). While there are many criticisms of the meme concept (e.g. Pinker 1997, see also Laland and Brown 2002), perhaps the most important is that memetics provides no greater insight into human behaviour than the more comprehensive and integrated theory of gene-culture coevolution (see below). However, while the meme concept may be subject to disinterest or skepticism within academia ,

the irony of its immense self-reflexive success in the public domain – notably to describe the viral spread of ideas, videos or other cultural snippets across the world wide web – is not lost on this author.

### 1.2.2 Gene-culture coevolution

Following a number of papers during the 1960s and 1970s (e.g. Campbell 1960, 1965, Feldman and Cavalli-Sforza 1976, Richerson and Boyd 1978), the field of gene-culture coevolution was first clearly defined in the 1980s with the publication of two hugely influential books: *Cultural transmission and evolution: a quantitative approach* (Cavalli-Sforza and Feldman 1981) and *Culture and the evolutionary process* (Boyd and Richerson 1985). These seminal works developed a number of mathematical models to describe the evolution of culturally transmitted traits and behaviours, and showed how the dual inheritance systems of genetics and culture can interact or coevolve. Drawing on models of population genetics, these gene-culture coevolutionary models showed how culture could be conceived of as the dynamic population-level result of processes of individual social transmission, in some ways analogous to – but in important ways different from – those of genetic evolution (see section 1.2.3).

Theoretical models developed by these authors, and subsequently many others (for example Aoki 1986, Durham 1991, 1992, Henrich 2004b, Feldman and Laland 1996, Laland *et al.* 2010, Richerson and Boyd 2005, Richerson *et al.* 2010), suggested that an evolving cultural system can, potentially rapidly, construct novel selection pressures on genetic loci. In contrast to memetics, models of gene-culture coevolution do not require *discrete* units of culture (see for example Henrich and Boyd 2002), though often in practice may make use of the meme-like

'cultural variant', which is then subjected to the forces of cultural evolution (see below).

This line of research has shed new light on long-standing human evolutionary questions, including the emergence of complex language (Aoki and Feldman 1987, Boyd and Richerson 2005, see also Deacon 1997), large-scale cooperative behaviour (Boyd and Richerson 1985, Henrich 2004b, Richerson and Boyd 2005, and see section 1.2.4) and the psychological mechanisms of social learning that themselves underlie the processes of cultural transmission (Boyd and Richerson 1985). Boyd and Richerson and colleagues (Boyd and Richerson 1985, Richerson *et al.* 2010) have argued that the evolution of such learning mechanisms may have been driven by the rapidly changing environmental conditions during the Pleistocene. The high-frequency high-amplitude climate fluctuations and subsequent ecological effects, would have placed a premium on the selective acquisition of up-to-date adaptive ecological information. One such mechanism, 'conformist transmission bias', in which the most common cultural trait or behavior is preferentially adopted, has been used to explain the evolution of pro-social or cooperative behaviour through the process of cultural group selection (Boyd and Richerson 1985, 2009, Henrich 2004b, and see section 1.2.4).

More recently in our evolutionary history, the cultural practice of farming, and the resulting dense populations and close contact with domesticates, provided novel selection pressures for genetic resistance to animal borne pathogens (Diamond 2002, Williamson *et al.* 2007). In addition, genetic adaptations associated with the consumption of novel agricultural and dairy products have been identified. Salivary amylase is an enzyme that digests starch, and a recent study (Perry *et al.* 2007) found evidence for positive selection on increased copy number in the salivary amylase gene (AMY1) in populations with a

history of high-starch diets typical of agriculturalists, compared to hunter-gatherer groups. Exceptionally strong signals of selection have been detected on a lactase (LCT) associated allele, which allows the digestion of the milk sugar lactose throughout adult life, in populations with a history of dairying (Bersaglieri *et al.* 2004, Ingram *et al.* 2009). The process by which the – now extremely common – lactase persistence allele spread throughout Neolithic Europe in tandem with the dissemination of dairying culture is the focus of Chapter 2 (see also Itan *et al.* 2009). Another striking adaptive genetic response to a cultural adaptation concerns the Kwa-speaking populations from west Africa, whose slash-and-burn agricultural practices inadvertently provided the ideal conditions for the proliferation of malarial mosquitoes. This created a novel and localized selection pressure on the HbS haemoglobin allele, which in homozygous individuals causes sickle-cell anaemia yet affords heterozygous individuals greater protection against malaria (Durham 1991, see also Laland *et al.* 2010).

Motivated by the mounting evidence of the impact of environmental modification – by both humans and other animals – on genetic selection pressures, Odling-Smee and colleagues have proposed the theory of 'niche construction' (Odling-Smee, Laland and Feldman 2003, Laland *et al.* 1999, 2001, Laland *et al.* 2010). They suggest augmenting the model of human dual-inheritance with a third line of ecological inheritance, transmitted alongside genetics and culture. Others authors do not always see the need for such an explicit division between ecological and cultural inheritance – viewing a culturally modified environment as simply another element of a broad cultural package (see Richerson *et al.* 2010). Niche construction studies are, in general, entirely consistent with those of gene-culture coevolution, and it remains to be seen

whether the additional conceptual detail pays its way by providing greater explanatory power.

The examples in this section should hopefully have made clear the critical importance of the interaction between culture and genetics in human evolutionary history, but what exactly does it mean to say that 'culture evolves'?

### 1.2.3 Culture as an evolutionary system

While there are clearly myriad definitions and interpretations of what gene-culture coevolutionists mean by the word 'culture', perhaps the most axiomatic is that of Boyd and Richerson (1985; page 33), who define it thus:

> *"Culture is information capable of affecting individuals' phenotypes which they acquire from other conspecifics by teaching or imitation."*

This definition stresses the importance of the social or 'cultural' transmission of information between individuals, clearly differentiating it from information obtained solely through individual learning or invention. Culture evolutionary theory draws on the 'population thinking' of genetics (Mayr 1982), seeing culture as a dynamic population-level result of such social transmission processes,

However, the mode of cultural transmission is more complex than that of genetics, as cultural information or traits can be transmitted in a number of ways other than the strict parent-child inheritance of genes. 'Naïve individuals' may acquire cultural traits from a genetic parent ('vertical transmission'), but also from non-parental individuals ('oblique transmission') or peers ('horizontal

transmission'), or a combination of any or all three of these modes. However, an evolutionary system requires more than just the transmission of information; evolution can perhaps be most parsimoniously described as *the differential success of inherited variation*. An evolutionary theory of culture must demonstrate that cultural variation is generated, and that there exist forces that affect the successful transmission of these cultural variants. So does the inheritance system of culture proposed by gene-culture coevolutionary theory possess these properties? Many authors (see reviews by Mesoudi *et al.* 2004, 2006 and Richerson and Boyd 2005) have demonstrated emphatically, in my view at least, that it does.

Cultural variation can arise through the transformation of cultural information during learning or transmission whether simply through random error, or by intentional modification, recombination or innovation ('guided variation') (see Mesoudi *et al.* 2004, and Chapter 3 for more detail). While there are obvious analogies between directionless copying error and random genetic mutation, the potential for making 'directional' or 'foresighted' improvements to cultural traits have led some authors to question the validity of evolutionary approaches to culture (Pinker 1997, but see Mesoudi 2008). However, as described in detail by Mesoudi *et al.* (2004), intellectual history is littered with accidental inventions and discoveries, some of them amongst the most ground-breaking, such as penicillin and X-rays. While this does not discount the fact that such directed modification allows cultural evolution to deviate from the tenets of genetic evolution, the minimal evolutionary requirement for variation to *exist* is fulfilled, and its mode of generation is less important than the action of cultural evolutionary forces upon it.

The final necessary condition of an evolutionary system is that of the differential success of such variation in a population; and in the case of cultural evolution this is driven by a number of mechanisms, some which have an analogue in genetics and some which do not. These are the so-called 'forces of cultural evolution', which include 'cultural drift', 'guided variation' (see above), and 'biased transmission' (Boyd and Richerson 1985).

Cultural drift is a direct analogue to genetic drift, and results in stochastic fluctuations in cultural variant frequencies over time simply as a result of sampling error in populations of finite size. In small populations such chance factors may result in drastic changes in frequency, or even loss, of a cultural trait. Cultural drift is covered in more detail in section 1.2.5, and forms the basis for the work presented in Chapters 4 and 5 of this thesis. Transmission biases occur when cultural variants are preferentially adopted during transmission, and serve to increase the frequency of such preferred variants. I devote the next section to the various cultural transmission biases that have been theoretically and empirically identified.

### 1.2.4 Cultural transmission biases

A distinction has been drawn between the 'content' and 'context' psychological biases that underlie social learning (Henrich and McElreath 2003; Box 3, page 129). This first category of biases, also called 'direct' bias (Boyd and Richerson 1985; page 135), involves individuals choosing an adaptive cultural trait either based on an intentional comparison – along the lines of cost-benefit analyses – between the various potential options, or 'innately' as a result of an evolved psychological preference (Henrich and McElreath 2008). Either way, the

preferential adoption, or otherwise, of a cultural trait depends on *its* intrinsic qualities.

By contrast, the second category of 'context' biases concerns the *source* of cultural information, and is *not necessarily* dependent on its content. Selecting a preferred cultural model to copy or adopting a trait based on its prevalence *may* provide a shortcut to gaining adaptive cultural traits that circumvents the potentially costly or time-consuming direct evaluation of the alternatives. Of course, whether cultural information gained in this way *is* adaptive depends on any number of factors; for example a rapidly changing environment may render this indirect information out of date. This issue forms the crux of much work on the evolution of social learning (see Rogers 1995, Boyd and Richerson 1995), and was one that the recent Social Learning Strategies Tournament was designed to address (Rendell *et al.* 2010). These context biases can be further divided into two useful subcategories, each with a distinct mechanism: 'model-based' biases and 'frequency-dependent' biases.

Model-based biases relate to the choice of an individual from whom to learn. Perhaps the most empirically interesting examples of this kind of bias are: 'success-bias' and ' prestige-bias'. Success-bias involves copiers choosing a model by making an evaluation, or ranking, of the (apparent) success of all potential cultural models within the cultural domain of interest. Observing cultural 'success' is an indirect means of measuring the skill level of a cultural model. By aggregating the results of a model's cultural behaviour over multiple 'performances', measures of success – such as the number of hunting kills – are often more reliable than direct observation of a small number of performances that are potentially subject to random chance events (see Henrich and McElreath 2008). However, prestige-bias causes copiers to preferentially choose a

particularly prestigious individual as a cultural model (Henrich and Gil-White 2001) for reasons not necessarily related to the cultural domain of interest. An individual may become prestigious for many reasons – including simply by reaching an old age (Henrich and Gil-White 2001) – regardless of their expertise or success (or lack thereof) within the focal cultural domain. Prestige-bias is a good, and perhaps the best, example of what Boyd and Richerson have called 'indirect-bias' (Boyd and Richerson 1985). A contemporary example would perhaps be the propensity for world leaders and international aid organizations to take advice on alleviating world poverty from celebrities, famous for their expertise in other areas, and the phenomenon of celebrity endorsement and advertising in general.

Frequency-dependent biases, however, are concerned with the relative frequency of a behaviour or trait in the population of potential cultural models; so preferential adoption is due to the trait being at high frequency, 'conformist' bias, or at low frequency, 'anti-conformist' (or novelty) bias. Some authors also use the equivalent terminology: 'positive' or 'negative' frequency-dependent bias, respectively (see Chapter 4; section 4.6).

In the absence of any observable correlation between success / prestige and cultural behaviour, i.e. an 'information-poor' situation (Henrich and McElreath 2003), a conformist transmission bias effectively allows a naïve individual to aggregate the behaviour of all potential models. If this population comprises a combination of individual learners (generating environment-specific adaptive cultural behaviour) and conformist social learners, then the frequency of a cultural trait acts as a cue of its adaptive quality. Theoretical modeling work has shown that adopting the majority cultural trait can be advantageous in heterogeneous environments (see Boyd and Richerson 1985; chapter 7, Henrich

and Boyd 1998), and, indeed, this kind of innate bias appears to evolve (given the presence of some individual learners) under a very wide range of conditions (Henrich and Boyd 1998). There is a wealth of empirical evidence for conformist cultural transmission (for an early example see Asch 1951), further supplemented by recent experimental laboratory work (Coultas 2004, McElreath *et al.* 2005).

Conformist bias is also the mechanism that some authors (for example see Boyd and Richerson 1985, 2009, Henrich 2004b, Richerson and Boyd 2005) have argued crucially underpins the evolution during the Late Pleistocene of the large-scale (or potentially parochial, e.g. Choi and Bowles 2007) altruistic or pro-social norms of behaviour that are so characteristic of our species. While space does not permit lengthy exposition of this argument, the basic line of reasoning is that conformism reduces within-group cultural variation, which serves to establish greater *inter*-group variation due to multiple stable 'cultural equilibria' being possible (Boyd and Richerson 2009). Cultural group-selection pressures, which may involve violence (Soltis *et al.* 1995, Choi and Bowles 2007, Lehman and Feldman 2008, Bowles 2009) or not (see Boyd and Richerson 2002, 2009), then acting on this inter-group level variation would tend to favour groups with more, or stronger, pro-social norms – which may include parochial warlike behaviour (Bowles 2009).

In another – no less important – context, conformist biases of this kind are also the mechanism proposed to explain the characteristic S-shaped cumulative adoption curves very extensively documented in the diffusion of innovations literature (Rogers 1995, see also Henrich 2001). However, a recent model (Kandler and Steele 2010) has shown that such curves can also be generated simply by assuming population heterogeneity in the ability to adopt a cultural innovation, with wealth inequality being one obvious example. The effects of

frequency-dependent biases are the focus of the work contained in Chapter 4; section 4.6.

### 1.2.5 Neutral, random-copying or drift models of cultural evolution

In contrast to models of non-random or biased acquisition of cultural information, some authors (Neiman 1995, Bentley *et al.* 2004, Shennan 2001, Bentley *et al.* 2009, Ormerod 2005) have found it fruitful to consider neutral models of random-copying derived directly from the neutral model of population genetics (Kimura and Crow 1964, Kimura 1983, see also Neiman 1995). In contexts where all possible variants of a cultural trait can be assumed to be selectively neutral, i.e. 'stylistic' rather than 'functional' (see Dunnell 1978, 1980, and Chapter 4 for further discussion), the evolution of such traits can be modeled as a process of cultural drift, where variant frequencies are subject to stochastic change simply through sampling error during transmission or copying. By direct analogy with the neutral genetic model, Neiman (1995) demonstrated that the amount of cultural variation expected at innovation-drift equilibrium would depend critically on the rate of innovation and the cultural population size (see Chapter 4; section 4.4.1).

It is worth clarifying here the difference between the neutral random-copying model and frequency-dependent biased copying. The neutral model, in which individuals copy the cultural variant of a *randomly* sampled individual, *results* in directly frequency-dependent variant frequency change, i.e. cultural drift. So there is no *intentional selection* of cultural model to copy. By contrast, frequency-dependent *copying biases* – such as conformism – are defined by the *non-random* choice of cultural model. A model is *selected* to copy based on the frequency of the variant they carry in the population. So while the variant a biased

copier obtains *is* frequency-dependent (but not necessarily in a linear way) from the viewpoint of the variants themselves, the important difference is that an *intentional choice* has been made regarding its acquisition.

I believe that there is room for both neutral and biased (or selective) models within cultural evolution, and that they provide complementary tools for investigating human culture, each within the domain to which they are best suited. Some authors have also drawn a distinction between what they call 'core' and 'peripheral' cultural traits (see for example Boyd *et al.* 1997). Core traits, which form the coherent basis of societal norms and traditions, are likely to be subject to selective transmission processes, such as conformity- and prestige-biases. Peripheral traits – some of which may be 'stylistic' (Dunnell 1978, 1980, see also Shennan 2008) – may be less tightly constrained, and may even be best described by stochastic cultural drift. The neutral or drift approach has been widely used as a simple null model of cultural change, and while no explicit psychological mechanism is posited it has provided what appeared to be remarkably good fits to some archaeological and cultural datasets (Neiman 1995, Shennan and Wilkinson 2001, Bentley *et al.* 2004, Schauer 2008, but see Chapter 4; section 4.5). This thesis contains work utilizing both neutral (Chapters 4 and 5) and biased (Chapter 3) models, as well as a model that combines elements of both approaches (Chapter 4; section 4.6, see also Mesoudi and Lycett 2009).

### 1.2.6 Gene-culture coevolution and cultural evolution: a summary

As has been outlined above, recent years have seen many researchers embracing evolutionary ideas and techniques across a wide range of disciplines, including archaeology (Neiman 1995, O'Brien and Lyman 2000, Shennan 2000,

2002, Lipo *et al.* 1997, Eerkens *et al.* 2006, Collard *et al.* 2006, Mesoudi and O'Brien 2008, Brantingham 2007), anthropology (Durham 1991, 1992, Mace and Pagel 1994, Holden and Mace 1997, Mace and Holden 2005, Tehrani and Collard 2002, 2009, Henrich and McElreath 2003, Henrich 2004a, Richerson and Boyd 2005, Borgerhoff Mulder *et al.* 2006, Greenhill *et al.* 2009), economics (Bowles 2004, Ormerod 1998, 2005, Beinhocker 2006, Gintis 2007), psychology (Mesoudi *et al.* 2004, 2006, Coultas 2004, McElreath *et al.* 2005) and linguistics (Gray and Jordan 2000, Gray and Atkinson 2003, Gray *et al.* 2009, Kandler 2009).

While many cultural evolution studies are based on models derived by analogy with population genetics as previously described, some use techniques developed in phylogenetics to investigate the evolutionary relationships of cultural artefacts or practices and languages (Lipo *et al.* 1997, O'Brien and Lyman 2000, Mace and Holden 2005, Gray and Jordan 2000, Gray *et al*. 2009). This sub-field of 'cultural or linguistic phylogenetics', uses observed contemporary variation to infer ancestral evolutionary lineages and derives phylogenetic relationships that provide evidence for past demographic processes. While there is an ongoing debate in genetics on the validity of inferring demographic history by phylogeographic means (Nielsen and Beaumont 2009), the case for cultural and linguistic phylogeographic inference seems on surer ground. In intra-species population genetics the non-equivalence between lineages and populations is clear (see Chapter 6; section 6.4.6), but by often treating the cultural or linguistic phylogenies as analogous to a species tree cultural evolutionary researchers can largely circumvent this problem. Indeed, while culture and language *are* population-level phenomena, the level of within-group conformity or homogeneity is far beyond what would ever be the case for genetics, making it

more likely that a reconstructed phylogeny would accurately recapitulate population history.

There is also a developing cross-fertilization between gene-culture coevolution and evolutionary game theory, championed by economists Herbert Gintis and Samuel Bowles and colleagues (Boyd *et al.* 2003, Bowles 2004, Henrich *et al.* 2005, Gintis 2007, 2009). In brief, game theory is the study of the strategic interaction of intentional agents, and adding an evolutionary dynamic allows the study of the emergence of adaptive strategies or behaviours (Maynard Smith 1982, Axelrod 1984). This work has played a central role in the development of theories of the evolution of cooperation and pro-social or altruistic behaviour (Axelrod 1984, Skyrms 1996, Nowak *et al.* 2004, Nowak 2006). Some authors (Laland and Brown 2002, Gintis 2007) have expressed the hope that an eventual synthesis between gene-culture coevolution and evolutionary game theory will provide a unifying theoretical framework for the study of human behaviour.

Much of the research carried out under the broad banner of gene-culture coevolution (and Chapters 3, 4 and 5 of this thesis) often in practice considers just cultural evolution with no genetic component, presupposing the existence of an evolved psychology equipped with the types of social or cultural learning biases previously outlined. This work both derives theoretical results on how factors such as population size and substructure can affect the evolutionary dynamics of cultural variants (e.g. Shennan 2001, Henrich 2004a, Strimling *et al.* 2009, Kandler and Laland 2009, Powell *et al.* 2009), and also uses empirically observed cultural variation to make inferences on population histories (e.g. Neiman 1995, Holden and Mace 1997, Gray and Jordan 2000, Shennan and Wilkinson 2001, Bentley *et al.* 2004). Naturally, general results arising from such cultural evolutionary

modelling can – if needed – be easily incorporated into gene-culture coevolution models developed in the broader parent discipline.

## 1.3 Population genetics and genetic variation

### 1.3.1 Human genetic diversity: a short history

The (indirect) study of human genetic diversity has a relatively long history; dating back to the very beginning of the 1900s with the (Nobel-prize-winning) discovery of the ABO blood group system (Landsteiner 1900), and the subsequent recognition that the frequencies of these inherited blood groups differ between populations. The development of protein electrophoresis in the mid-1900s allowed the detection of polymorphisms in blood proteins, including haemoglobin, a variety of immunoglobulins and the human leukocyte antigen (HLA) system – which is highly polymorphic. A famous early analysis of a subset of these indirect 'classical markers', sampled globally from seven broad 'racial' groups, showed that the majority of diversity (~85.4%) was mainly *within* populations rather than *between* them (Lewontin 1972, but see also Edwards 2003).

Drawing together these large datasets, the seminal work of Cavalli-Sforza and colleagues (for example Menozzi *et al.* 1978, Cavalli-Sforza *et al.* 1994) analyzed the worldwide variation at multiple loci simultaneously. Using principal component analysis (PCA) – a technique for summarizing high-dimensional data while retaining explanatory power – these early studies showed a degree of (broadly) continental population differentiation, and produced synthetic maps (see Cavalli-Sforza *et al.* 1994) that allowed the visualization of the geographic structure of worldwide genetic variation for the first time. The clinal patterns of

the principal components (PC) were interpreted, in some instances, as evidence for past demographic events, such as the initial colonization of the Old World from Africa (e.g. Cavalli-Sforza 1994; page 135) or the spread of Neolithic farmers from the Near East into Europe (*ibid*.; page 292). However, recent spatially-explicit modeling work has shown that such direct interpretations may be problematic (more below, and see Novembre and Stephens 2008).

Subsequent developments – specifically the invention of the polymerase chain reaction (PCR) (Saiki *et al.* 1985, Mullis and Faloona 1987) – allowed the direct analysis of DNA sequence data, rather than the indirect means afforded by the previous classical markers. The first phylogenetic analyses of mitochondrial DNA (mtDNA) (e.g. Cann *et al.* 1987, Vigilant *et al.* 1991) indicated high African diversity compared to non-African populations, and a recent African common maternal ancestor ~160 kya – the highly publicized, and much misunderstood, 'mitochondrial Eve'. Subsequent studies on autosomal markers – such as restriction fragment length polymorphisms (RFLP) and microsatellites (Nei and Roychoudhury 1993, Bowcock *et al.* 1994, Tishkoff *et al.* 1996, see also Jorde *et al.* 2000) – and on Y-chromosome data (e.g. Donnelly *et al.* 1996, Pritchard *et al.* 1999) provided further evidence for a recent common ancestor. These and other studies showed that, while human genetic diversity is low compared to other great ape species (Kaessmann 2001), genetic variation generally appears to be greatest in sub-Saharan Africa, and that non-African populations carry just a subset of this African diversity. Taken together, this work provided strong support for the Recent African Origin model. The relative lack of diversity observed in non-African populations was seen as evidence of a recent, and potentially severe (Harpending *et al.* 1993, Sherry *et al.* 1997), population bottleneck prior to the population expansion from the continent that colonized

the rest of the world (e.g. Harpending *et al.* 1993, Sherry *et al.* 1994, Rogers 1995).

### 1.3.2 Human genetic diversity in the genomic era

The transition into the 'genomic era' – perhaps most prominently heralded by the highly-publicized Human Genome Project (Lander *et al.* 2001, Venter *et al.* 2001) – has dramatically increased the availability of single nucleotide polymorphism (SNP) and DNA sequence data. In tandem with the necessary, and rapidly developing, statistical and computational analysis methods (e.g. Beaumont and Rannala 2004, Marjoram and Tavaré 2006), these data have allowed for a greater understanding of the selective and demographic processes that shaped the human genome. While the Human Genome Project provided a draft of a single 'reference' genome, ongoing large-scale collaborative projects, such as the Human Genome Diversity Project (HGDP) (Cavalli-Sforza 2005) and the International HapMap Project (The International HapMap Consortium 2007), are attempting to comprehensively map the extent of world-wide human genetic diversity. There are potential benefits for improving the understanding of genetic diseases and for pharmacogenetics, which studies the genetic basis of variation in drug metabolism in order to better inform drug administration policy and reduce the risk of adverse reactions. These datasets are also proving invaluable to geneticists – and genetically-inclined anthropologists and archaeologists – interested in uncovering past population dynamics to shed light on the evolutionary history of our species.

In line with the previous studies outlined above, analyses using this genome-wide data found that at the global-scale human genetic diversity – sampled from 52 worldwide populations – clusters similarly into 5 broad

continental groups: sub-Saharan Africa, west Eurasia, east Asia, Oceania and the Americas (Rosenberg *et al.* 2002, Li *et al.* 2008). As predicted by the early theoretical genetic models of 'isolation-by-distance' (Wright 1943, Malécot 1948) or the 'stepping-stone' model (Kimura and Weiss 1964), analyses of this data revealed that genetic differentiation between populations increases with distance (Ramachandran *et al.* 2005, Prugnolle *et al.* 2005, Rosenberg *et al.* 2002, Handley *et al*. 2007). These studies found that global genetic diversity decreases almost linearly with increasing geographic distance from east Africa, and was best explained by a serial founder model of expansion from this region (see also DeGiorgio *et al.* 2009). A recent study using African HGDP data (Tishkoff *et al.* 2009) provided a detailed picture of the continental structure of genetic variation, and confirmed the previous findings of high genetic diversity in sub-Saharan Africa. Cumulatively this work provides overwhelming evidence in support of the RAO model, though – as mentioned above – the indication of some level of interspecies gene flow between modern humans and Neandertals (Green *et al.* 2010) means that the idea of strict replacement of Neandertals by *Homo sapiens* will need to be slightly revised (see also Fagundes *et al.* 2007).

### 1.3.3 Disentangling demography and selection

The global geographic distribution of genetic diversity so far outlined has been largely explained in terms of past demographic processes; where population size and substructure causes inter-population differences in allele frequencies mainly through bottlenecks and drift. However, there is accumulating evidence for the effects of natural selection acting on various genetic loci. Some of the selective pressures appear to be shaped by our expansion into new ecological

environments and niches, with perhaps the most obvious examples being on loci associated with readily-observable phenotypic differences such as skin pigmentation and hair morphology (Voight *et al.* 2006, Williamson *et al.* 2007, and see also Jablonski and Chaplin 2000). Other loci appear to be subject to gene-culture coevolutionary selective pressures, as previously discussed in section 1.2.2.

Detecting the signals of natural selection from modern genetic data can be a difficult undertaking, as they will – to varying extents – be disguised by past demographic processes (Novembre and Di Rienzo 2009). Regional differences in allele frequencies may be driven purely by drift in structured populations with restricted gene flow (see below). So while brute force genome scans may identify many candidate loci, some may have a more parsimonious demographic explanation (but see also Williamson *et al.* 2007). Without a knowledge of the geographical distribution of genetic variation expected in the *absence* of selection it can be difficult to accurately identify those alleles that have been subject to selective pressures (Coop *et al.* 2009). Major recent advances in population genetics, such as coalescent-based simulation approaches (e.g. Hudson 1991, see Chapter 6; section 6.4 for more detail), have allowed for the explicit consideration of increasingly complex demographic histories (for example Ray *et al*. 2005, Fagundes *et al.* 2007). Genetic models incorporating demography, which may be conditioned with independent lines of evidence from palaeoanthropology and archaeology, allow the relatively straightforward generation of simulated neutral genetic data that approximates what we should expect to see in the absence of selection. Comparing this neutral, or null, expectation of the geographic structure of genetic variation with that we actually observe may allow the identification of candidate loci that have undergone selection (e.g. Akey 2002).

Perhaps one of the most interesting, and important, phenomena that this spatially explicit demographic modeling approach has uncovered is that of 'allele surfing'. Recent simulation work (Edmonds *et al.* 2004, Klopfstein *et al.* 2006) – and subsequent work on real bacterial populations (Hallatschek and Nelson 2009) – has demonstrated that when a population is undergoing a spatial expansion, alleles can 'surf' to high frequency across a large region on the front of the wave of expansion. This is due to the magnified effects of drift on allele frequencies in the relatively small 'pioneer' populations, effectively formed through sequential sampling of the larger populations on or near the expanding wavefront. This surfing effect can act either on existing, or 'standing', variation in the population or, perhaps more strikingly, can rapidly propagate a novel mutation that arises on or near the wavefront over a wide geographical range and carry it to a high frequency – even if deleterious (for example Hallatschek and Nelson 2009). The spatial distribution of an allele following an expansion can thus have a centroid radically displaced from its origin (Klopfstein *et al.* 2006; figure 1A). Without explicit consideration of the possible causative demographic process, interpreting spatial patterns of allele frequencies may result in misleading inferences about that allele's history, or even require that selective pressures be invoked in order to explain the distribution (see Currat *et al.* 2006).

The effects of a colonizing expansion into a previously unoccupied region have long been considered, and models discussed above (e.g. Ramachandran *et al.* 2005, DeGiorgio *et al.* 2009) show the impact of sequential or 'serial' drift at the wavefront of expansion from Africa (see also Novembre and Di Rienzo 2009). These 'serial founder' models of human expansion can give rise to allele surfing, which has been proposed to explain regional differences in allele frequencies for which localized selection pressures had previously been invoked, e.g. the neural-

related genes ASPM and microcephalin (Mekel-Bobrov *et al.* 2005, Evans *et al.* 2005, and reply from Currat *et al.* 2006, see also Voight *et al.* 2006 and Williamson *et al.* 2007 for lack of evidence for selection). Indeed, Hofer *et al.* (2009) have demonstrated that many large allele frequency differences between continents – from global HDGP data – can be explained by demographic range expansions rather than by localised selection pressures.

However, a population expansion into an *already occupied region*, such as occurred – to some extent – in the Neolithic expansion of farming into Palaeolithic Europe (see Chapter 2; section 2.3), may not be as straightforward due to allele surfing (see also Currat and Excoffier 2005). Perhaps contrary to intuition, even very minimal levels of gene flow between a low-density indigenous population and a high-density incoming population can be 'massively' magnified by the surfing phenomenon (Currat *et al.* 2008). This 'introgression' means that the true extent of the colonization may be disguised, as the colonizing pioneers can end up carrying a large proportion of indigenous genes (see also Green *et al.* 2010 on admixture between *Homo sapiens* and Neandertals).

As mentioned above, a recent study by Novembre and Stephens (2008) has shown that interpreting the results of PC maps as direct evidence for past demographic expansions may be unjustified. The authors performed PCA on simulated genetic data generated under null equilibrium models, i.e. with no range expansions, that allowed homogenous short-range migration. They found that the first 4 PCs showed the same large-scale patterns (Novembre and Stephens 2008; figure 1) as many of the continental PC maps presented by Cavalli-Sforza and colleagues (Cavalli-Sforza *et al*. 1994), meaning that these gradients could have been generated purely through differentiation due to geographic distance rather than expansion. Another recent study, by François *et al.* (2010),

has also demonstrated, contrary to previous expectation, that a population expansion can very often result in a spatial distribution where the greatest gradient of genetic variation, i.e. the first principal component PC1, is perpendicular to the direction of expansion, rather than parallel to it (see Cavalli-Sforza *et al.* 1994). They show that this is due to allele surfing creating potentially multiple different 'sectors' of allelic homogeneity, or even fixation, which propagate outwards in the direction of the expansion (François *et al.* 2010; figure 6). Following the expansion, the greatest gradient of variation (PC1) is often in a direction which directly cuts *across* these multiple genetically differentiated (yet internally homogeneous) sectors, i.e. perpendicular to the expansion direction.

In summary, it is clear that demography is exceptionally important in attempting to interpret observed geographical patterns of genetic variation. Understanding the diversity expected under neutrality can help identify loci that have been, or are still, under selection, and also in inferring the demographic history of a population. While this last point may seem circular, it is only by explicitly including demographic processes in model-based analyses that evaluation of the specific competing hypotheses about population history can be made (Nielsen and Beaumont 2009, see also Chapter 6; section 6.4.6). Coalescent theory, and its subsequent implementation in efficiently generating simulated genetic data, provides a powerful means to test increasingly complex past demographic scenarios with increasingly large and fine-grained genetic datasets. More detail is provided in Chapter 6, and an overview of recent advances in associated statistical methods, including approximate Bayesian computation (ABC), is given in Chapters 2 and 6.

## 1.4 Demography and the evolution of genetic and cultural variation

While in population genetics the importance of population size and substructure has long been obvious, the effects of demography on cultural evolution are only just starting to be formally addressed (Boyd and Richerson 1985, Cavalli-Sforza and Feldman 1981, Neiman 1995, Shennan 2001, Henrich 2004a, Bentley *et al.* 2004, Ghirlanda and Enquist 2007, Powell *et al.* 2009, Kline and Boyd 2010). The idea of effective population size $N_e$ in population genetics was introduced by Wright (1931), and provided a way to rescale a real population to an idealized Wright-Fisher population subject to the same level of genetic drift. While there are a number of ways, not necessarily equivalent, of defining effective population size, based on variance in allele frequencies, inbreeding and coalescence (see Ewens 2004, Sjödin *et al.* 2005); in general the effective size is substantially smaller than the population census size.

Models of neutral cultural evolution derived from population genetics (see for example Neiman 1995, and Chapters 4 and 5 of this thesis), analogously use the idea of a cultural effective population size. This can be thought of as either the size of the population of encultured individuals or the size of the population of artefacts, or 'cultural pool', available as cultural models for naïve learners. Many archaeological studies applying the neutral cultural model (Neiman 1995, Shennan and Wilkinson 2001, Bentley *et al.* 2004, see also Chapter 5) indicate that effective population size, and substructure (Neiman 1995), may determine the level of cultural variation that can be maintained within a population. This work is fully reviewed in Chapter 4.

Other cultural evolutionary models (Shennan 2001, Henrich 2004a), while invoking different cultural transmission mechanisms – drift and direct bias respectively – have demonstrated the critical importance of effective cultural population size on the accumulation of culturally inherited traits. Building on the model of Henrich (2004a), Chapter 3 also demonstrates the importance of population substructure, showing that spatial variation in cultural complexity can arise in regions of heterogeneous subpopulation density. This result was used, in conjunction with archaeological and genetic evidence, to argue that the temporal and geographical heterogeneity in the earliest appearance of modern human behaviour (McBrearty and Brooks 2000) was caused by heterogeneity in effective cultural population size (Powell *et al.* 2009, see also Shennan 2001, Stringer 2007). Recently Kline and Boyd (2010) have provided further support for Henrich's general model, showing that the technological toolkit complexity of Oceanian islanders at the time of European contact was correlated with population size and the degree of inter-island contact.

As outlined in the previous sections, the emergence of novel cultural practices that affect population size, structure and density, such as farming, has in some cases led to changes in selection pressures on various genetic loci (Hawks *et al.* 2007, Voight *et al.* 2006, and see Laland *et al.* 2010 for a recent review). These two coevolving inheritance systems can give rise to otherwise unexpected geographic distributions of alleles. Comprehensive explanations for the observed variation in human genetic and cultural systems will require a gene-culture coevolutionary approach, and increasing our understanding of the way demography can affect both lines of inheritance will be crucial in this endeavour. The following thesis presents my work on a wide range of topics related by this common thread.

Chapter 2 presents a simulation model used to investigate the gene-culture coevolution of lactase persistence and dairy farming in Neolithic Europe (published as the co-authored paper Itan *et al.* 2009). We consider 3 distinct populations which interact both genetically and culturally: indigenous hunter-gatherers, non-dairying farmers and dairying farmers, and model the spread of a lactase persistence allele across Europe, selected for only in dairy farming populations. By explicitly modeling the spatial expansion of farmers from the Near East into Europe, this model also incorporates the phenomenon of allele surfing discussed in the previous section. To analyze this model I used approximate Bayesian computation (ABC), which allowed the integration of both modern genetic data on the European lactase persistence-associated allele and archaeological data on the arrival of farming across Europe. I provide full details on the background and development of this statistical technique, and it is used further in Chapter 6. We inferred a time and geographic region for the origin of this coevolutionary process that fits well with the emergence in central Europe of the Linearbandkeramik (LBK), one of the earliest dairying cultures.

Chapter 3, published as Powell *et al.* (2009), develops further the model of Henrich (2004a), which described the effect of population size on the accumulation of culturally inherited skills. We placed our extended model in simulated conditions approximating the demography history of our species during the Late Pleistocene, and showed that geographic heterogeneity in subpopulation density can lead to stable spatial structuring in cultural accumulation. The appearance of modern human behaviour during this period is most often associated with the Upper Palaeolithic / Late Stone Age (Mellars 2005, Bar-Yosef 2002), and has long been explained as a result of neural- or cognition-related genetic change (see for example Klein 1999, 2000, Mithen 1996).

However, such arguments do not satisfactorily account for the increasing amount of evidence for the sporadic appearance of markers of behavioral modernity at far earlier Middle Stone Age sites across sub-Saharan Africa (McBrearty and Brooks 2000, Henshilwood *et al.* 2002, Marean *et al.* 2007, Henshilwood *et al.* 2009). Combining the results of our model with previously published genetic estimates of regional effective population size through time (Atkinson *et al.* 2008), we demonstrate that population density is a critical factor in explaining the temporally and geographically heterogeneous appearance of modern human behaviour in the archaeological record.

The next two chapters (Chapters 4 and 5) are related in that they are both concern neutral evolutionary models of cultural change. Chapter 4 provides a derivation of the neutral model and an overview of its application in archaeology (see for example Neiman 1995, Bentley *et al.* 2004). I demonstrate the need for a statistical test to detect deviation from neutrality in a cultural dataset. I then develop such a test and discuss its application to a large dataset of decorated pottery from Classical-era Athens (Schauer 2008). This test is then applied to models that allow a degree of (non-neutral) frequency dependent-copying bias (Mesoudi and Lycett 2009). The results show that surprisingly high levels of non-neutral copying can occur without deviation from neutrality being detected. Chapter 5 develops a novel neutral model of cultural evolution based on the Moran model of population genetics (Moran 1958, 1962). This model relaxes two of the major assumptions made in previous cultural evolutionary models; namely those of 'cultural equilibrium' and the idea of discrete 'cultural generations' (see for example Neiman 1995). Using an approximate maximum likelihood method – which relies on Monte Carlo simulation – this model allows the estimation of the effective population size of cultural artefacts and their rate of turnover from

minimal archaeological data in the form of frequency seriation, or 'battleship curve' peaks. Testing of this model on simulated datasets shows that accurate estimation of these parameters is possible using as few as 10 such data points, and I give an example of its application to a dataset of decorated LBK pottery from southwestern Germany.

Finally, in Chapter 6 I outline the coalescent model of population genetics, with a focus on its use in inferring past demography from both modern and ancient DNA samples. The case for population genetic modeling under an explicit demographic model as opposed to phylogeographic approaches is discussed. I then give two examples of my work using coalescent simulation coupled with ABC, which form part of two collaborative projects on inferring the demographic history of the Hispaniolan hutia (Brace *et al.* (in prep)) and investigating the domestication of Near Eastern taurine cattle (Bollongino *et al.* (under review)).

# Chapter 2

# Lactase persistence and dairying in Europe: a gene-culture coevolutionary model[1]

## 2.1 Summary

Lactase persistence (LP) is common among people of European ancestry, but with the exception of some African, Middle-Eastern and southern Asian groups, is rare or absent elsewhere in the world. A recently derived allele, -13,910*T, is strongly associated with LP in Europeans, and has been subject to strong positive selection. Further, ancient DNA work has shown that the -13,910*T allele was very rare or absent in early Neolithic central Europeans. LP would likely provide a selective advantage only with a supply of fresh milk, and this has led to the development of a gene-culture coevolutionary model where lactase persistence is favoured only in cultures practicing dairying.

I present a collaboratively developed demic computer simulation model that explores the spread of LP and dairying across Europe. Using both genetic and archaeological data we infer that the -13,910*T allele first underwent selection

[1] Published as Itan Y., Powell A., Beaumont M.A., Burger J., Thomas M.G. (2009) 'The origins of lactase persistence in Europe' *PLoS Computation Biology* 5(8): e1000491.

among dairying farmers around 7,500 years ago in a region between the central Balkans and central Europe, possibly in association with the dissemination of the Neolithic Linearbandkeramik (LBK) culture over central Europe. Furthermore, our results suggest that natural selection favouring an LP allele need not be higher in northern latitudes, as assumed by the *calcium-assimilation* hypothesis. Our results provide a coherent and spatially explicit picture of the coevolution of LP and dairying in Europe.

## 2.2 Abstract

I first outline our gene-culture coevolutionary model, which simulates the demographic and genetic history of three distinct but interacting cultural populations: hunter-gatherers, non-dairying farmers and dairying farmers. The model incorporates a number of demographic, genetic and cultural processes, and tracks the spread of a lactase persistence (LP) allele which emerges stochastically as the farming groups expand into Europe from their origin in Anatolia some ~10, 500 years ago, and is subject to selection solely in the dairying farmer population.

I then provide a detailed review of statistical inference methods, with a focus on approximate Bayesian computation (ABC), before detailing the ABC analysis I employed to estimate the parameters of interest from our relatively complex model. This analysis combines two lines of evidence; genetic data, in the form of -13,910*T allele frequencies in present day Europeans, and archaeological data, in the form of carbon-14 dates for the earliest arrival of farming across Europe.

## 2.3 Introduction

Lactase persistence (LP) is an autosomal dominant trait enabling the continued production of the enzyme lactase throughout adult life. Production of lactase in the gut is essential for the digestion of the milk sugar lactose. Lactase non-persistence – where the production of lactase ceases after weaning – is the ancestral condition for humans, and indeed for all mammals (Swallow 2003). However, LP is common in northern and western Europeans as well as in many African, Middle Eastern and southern Asian pastoralist groups, remaining rare or absent elsewhere in the world (Swallow 2003, Ingram *et al.* 2007, Mulcare *et al.* 2004, Tishkoff *et al.* 2007). In Europeans the LP trait is strongly associated with a single nucleotide polymorphism (SNP) -13,910*T, located 13.91 kb upstream from the lactase gene (LCT) in an intron of the neighbouring MCM6 gene (Enattah *et al.* 2002). *In vitro* studies (Olds and Sibley 2003, Lewinksy *et al.* 2005) have shown that this -13,910*T allele can have a direct effect on LCT gene promoter activity.

The -13,910*T allele ranges in frequency from 6–36% in eastern and southern Europe, 56–67% in central and western Europe, to 73–95% in the British Isles and Scandinavia (Mulcare 2006, Bersaglieri *et al.* 2004), while LP ranges in frequency from 15–54% in eastern and southern Europe, 62–86% in central and western Europe, to 89–96% in the British Isles and Scandinavia (Ingram *et al.* 2009). This spatial distribution of -13,910*T frequencies (see also table 2.2, and for example Ingram *et al.* 2009, Itan *et al.* 2010) makes this allele a good candidate for predicting LP in Europe. However, the -13,910*T allele cannot account for LP frequencies in most African and middle Eastern populations (Mulcare *et al.* 2004), where other

independent but similarly located LP-associated alleles have been identified (Ingram *et al.* 2007, Tishkoff *et al.* 2007, Enattah *et al.* 2007, Enattah *et al.* 2008), indicating convergent evolution.

The age of the European -13,910*T allele has been estimated by considering the modern genetic variation in the regions surrounding the LCT gene. Using long-range haplotype conservation Bersaglieri *et al.* (2004) estimated an age of between 2,188 and 20,650 years ago (ya), while Coelho *et al.* (2005) estimated between 7,450 and 12,300 ya using the variation at closely linked microsatellites. These relatively recent age estimates for an allele that is at such high frequency in parts of present day Europe indicate that -13,910*T has been subject to very strong selection. Indeed, the very large estimated selection strength 1.4–19% (Bersaglieri *et al.* 2004) is among the highest estimated for any human allele in the last ~30,000 years (Sabeti *et al.* 2006). Interestingly, similar age (1,200 to 23,200 ya) and selection strength (1–15%) estimates were also obtained for one of the major African LP-associated variants -14,010*C (Tishkoff *et al.* 2007).

It is unlikely that LP would provide any selective advantage without the regular supply of fresh milk, which would be provided only in a dairying culture. This has led to the development of gene-culture coevolutionary models; the *culture-historical* hypothesis where lactase persistence is only favoured in cultures practicing dairying (Kretchmer 1972, Simoons 1970, McCracken 1971a, Aoki 1986); and the *reverse-cause* hypothesis in which dairying is more favoured in LP populations (Simoons 1970, Bayless *et al*. 1971, Nei and Saitou 1986, McCracken 1971b).

However, the reasons why LP, in conjunction with dairying, should confer *such* a strong selective advantage remain open to speculation. Flatz and Rotthauwe (1973) proposed the *calcium assimilation hypothesis*, where an LP allele is favoured in high-latitude regions because reduced levels of sunlight do not allow sufficient synthesis of vitamin D in the skin. Vitamin D is required for calcium absorption and milk provides a good dietary source of both nutrients. This hypothesis, which suggests latitude-dependent selection strength, is consistent with the observed geographic distribution of LP and the -13,910*T allele. Other explanations for the selection of LP include: the ability to consume a calorie and protein-rich food source, the relative constancy in the supply of milk (in contrast to the boom-and-bust of seasonal crops), and the value of fresh milk as a source of uncontaminated fluids. It is likely that the relative advantages conferred by these various factors differ in Europe and Africa.

Estimates of the age of the -13,910*T allele correspond well with estimates of the onset of dairying in Europe. Slaughtering age profiles in sheep, goats and cattle suggest dairying was present in South-eastern Europe at the onset of the Neolithic (Vigne and Helmer 2007, Bartosiewicz 2007), while residual milk proteins preserved in ceramic vessels provide evidence for dairying in present day Romania and Hungary 7,900–7,450 ya (Craig *et al.* 2005). Furthermore, residual analyses of fats indicate dairying at the onset of the Neolithic in England, some 6,100 ya (Copley *et al.* 2003, Copley *et al.* 2005), and after 8,500 ya in the western parts of present day Turkey (Evershed *et al.* 2008). Allelic age estimates are also consistent with the results of a recent ancient DNA study (Burger *et al.* 2007) which showed that the -13,910*T allele was rare or absent among early farmers from Central and Eastern Europe. These observations lend support to the view that -13,910*T, and thus LP, rose rapidly in

frequency only after the onset of dairying, as opposed to the 'reverse-cause' hypothesis (Simoons 1970, Bayless *et al*. 1971, Nei and Saitou 1986, McCracken 1971b), whereby dairying developed in response to the evolution of LP.

Important questions remain regarding the location of the earliest -13,910*T-carrying dairy farming groups and the demographic and gene-culture coevolutionary processes that shaped the modern distribution of LP in Europe. The present day distribution of the -13,910*T allele (see table 2.2 and Itan *et al.* 2010 for example) might suggest an origin in northwestern Europe. However, this would seem to contradict the fact that the earliest evidence of dairying – in the form of archaeozoological remains, and lipid and protein residues – is found in the near East and southeast and Mediterranean Europe (Vigne and Helmer 2007, Evershed *et al*. 2008, Vigne 2006). Recent simulation modelling work (Edmonds *et al.* 2004, Klofpstein *et al.* 2006) has demonstrated that when a population is expanding spatially, an allele can be carried to high frequency on the front of the wave of expansion, a phenomenon called 'allele surfing'. This is due to the relatively low population densities at or near the wavefront, meaning that the effects of drift will be magnified as expansion occurs through sequential sampling of these small populations. This can result in the centroid of the spatial distribution of an allele (Klopfstein *et al.* 2006; figure 1A) being far removed from its origin, and can confound inference about its history unless this demographic process is explicitly considered (see Chapter 1; section 1.2).

The nature of the spread of farming into Europe during the Neolithic, starting ~10, 500 ya, has long been debated; was it a demic expansion of farmers themselves, or the cultural diffusion of farming technology and practices? (see for example Childe 1942, Clark 1965, Zvelebil and Zvelebil 1988) A number of previous genetic studies (Menozzi *et al.* 1978, Ammerman and Cavalli-Sforza 1984, Cavalli-Sforza *et al.* 1994, Sokal *et al.* 1991, Chikhi *et al.* 2002, Currat and Excoffier 2005) have attempted to quantify the proportion of contribution to the modern European gene pool made by the pre-existing Palaeolithic hunter-gatherer and the incoming Neolithic farmers. The clinal patterns of genetic variation produced through principal component analysis (PCA) demonstrated by Cavalli-Sforza and colleagues (see for example Menozzi *et al.* 1978, Ammerman and Cavalli-Sforza 1984) were widely interpreted as indicating a demic expansion from the southeast to northwest, with little Palaeolithic admixture. However, the southeast-northwest European PCA cline could simply be an artefact of the initial Palaeolithic expansion (Zvelebil and Zvelebil 1988), and recent work (François *et al.* 2010, and see also Novembre and Stephens 2008) has shown that the gradient of greatest genetic variation following a range expansion may in fact be perpendicular to the direction of expansion, due to the allele surfing phenomenon.

A number of other studies, using: Y-chromosome (e.g. Semino *et al*. 2000, but see also Chikhi *et al.* 2002 ), mitochondrial (mtDNA) (e.g. Richards *et al.* 2000), and autosomal (Chikhi *et al.* 1998, Dupanloup *et al.* 2004) DNA data, have indicated Neolithic genetic contributions of anywhere between 15 and 65%. Currat and Excoffier (2005) developed a semi-realistic stochastic forward simulation model, which allowed both gene flow and cultural diffusion between the population of Paleolithic hunter-gatherers and colonizing Neolithic farmers. They found that even

minimal levels of inter-group gene flow led to a large contribution of Palaeolithic lineages to the modern gene pool. So while there is still no general consensus, it appears unlikely that the Neolithic spread across Europe through a process of pure demic diffusion, and so cultural processes may be equally, if not more, important.

As a gene-culture coevolutionary process, it is clear that any explanatory model for the spread of LP in Europe must incorporate both cultural and genetic mechanisms. So in order to examine the potential scenarios, we developed a demic forward computer simulation model, which incorporated three distinct interacting cultural groups: hunter-gatherers, non-dairying farmers and dairying farmers. These groups underwent a number of genetic and demographic processes, both within- and between- groups, which simulated the spread of farming and farmers' genes across Europe. We tracked the spread of an allele that initially occurs near the wavefront of a farming population expansion and that was only under selection in the dairying farmer group. We compared the simulated predicted frequencies of this LP allele and arrival dates of farmers to known frequencies of the -13,910*T allele (Mulcare *et al.* 2004, Bersaglieri *et al.* 2004) and carbon-14 based estimates of the arrival dates of farmers (Pinhasi *et al.* 2005) at different locations throughout Europe.

We employed approximate Bayesian computation (ABC), a method that allows the estimation of parameters under models too complex for a full-likelihood approach (see section 2.5). By comparing summary statistics on the observed archaeological and genetic data with those computed on our simulated datasets, ABC enabled us to estimate the key demographic and evolutionary parameters including the region where LP-dairying coevolution in began in Europe.

## 2.4 The gene-culture coevolution simulation model

The gene-culture coevolutionary simulation model I describe here was created by Yuval Itan, Mark Thomas and myself in order of contribution (see Itan *et al.* 2009). It was written in the Python programming language (van Rossum *et al.* URL: http://www.python.org).

### 2.4.1 The simulation world

Our model divides the west Eurasian landmass into a collection of 2,375 land demes (see figure 2.1), each of which has a number of attributes that determine its demographic capacity. Each deme has an elevation factor *e*, calculated as 1 minus its elevation relative to the maximum (see Itan *et al.* 2009; figure S1a), and a climate factor *c*, which is given for 3 broad climate zones (see Itan *et al.* 2009; figure S1b):

$$c = \begin{cases} 1/3 & \text{cold/desert} \\ 2/3 & \text{temperate} \\ 1 & \text{Mediterranean} \end{cases}$$

The area of each deme in km$^2$ $A_{\text{deme}}$ is calculated to take into account the curvature of the Earth (see Itan *et al.* 2009). We assumed a 1:4 ratio for the relative size of the effect of climate and elevation on carrying capacities. Given this the total carrying capacity of each deme is given by:

$$K_{\text{deme}} = (0.2c + 0.8e)D_{\max}A_{\text{deme}}$$

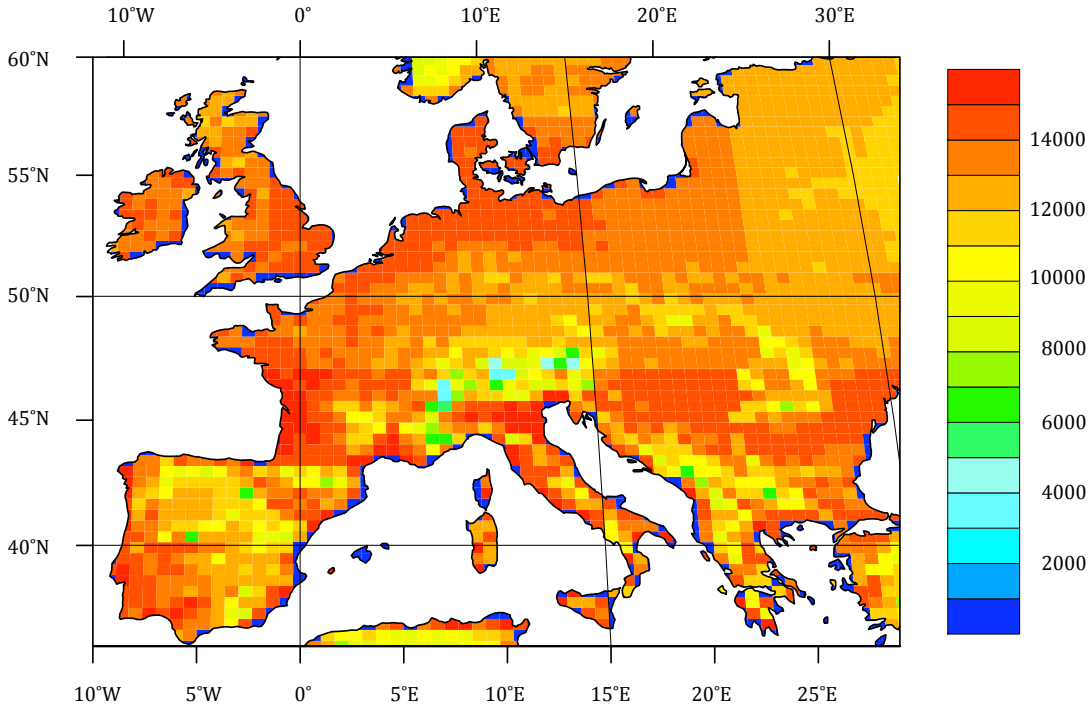
where $D_{max}$ is the maximum population density fixed at 5 individuals per km$^2$ (Hassan 1981). This scaling of carrying capacity is intuitive, in that a low-elevation Mediterranean deme will support a larger population than an equally sized high-elevation tundra or desert deme. Figure 2.1 shows the simulation world of demes with total carrying capacities $K_{deme}$. Elevation also has an effect on the distance that individuals can migrate; allowing individuals greater mobility at lower elevations. A relative mobility factor $M_{curr}$ is determined for each deme as:

$$M_{curr} = \begin{cases} 0.5 & \text{mountainous } > 1100 \text{ m} \\ 1 & \text{lowlands } \leq 1100 \text{ m} \\ 1.5 & \text{coastal demes} \end{cases}$$

Within each deme we can potentially model three distinct cultural groups: hunter-gatherers *HG*, non-dairying farmers $F_{ND}$ and dairying farmers $F_D$. As a proportion of the total carrying capacity $K_{deme}$ the relative carrying capacities of each cultural group are given by the ratio HG: $F_{ND}$: $F_D$ are 1:50:50, consistent with population densities estimates for farmers and hunter-gatherers (Bellwood 2005, Hassan 1981).

We simulated the spread of farming into Europe under a number of demographic, genetic and cultural processes. We tracked the frequency of an LP allele subject to selection *only* in the $F_D$ group, and also modelled a 'genetic background allele', which allowed an estimation of the contribution of the initial LP-dairying gene-culture coevolutionary group to the modern European gene pool.

**Figure 2.1:** Map of the simulation space, showing the carrying capacities of each deme $K_{\text{deme}}$, which are determined by elevation, climate and area (see text).




### 2.4.2 Demographic processes

Each new simulation of the model starts with *HG* populating every deme, and a population of each farming group $F_{\text{ND}}$ and $F_{\text{D}}$ in Anatolia. We allow all three populations to grow and migrate, simulating the spread of both farming groups into Europe.

Each cultural group in each deme undergoes logistic growth every generation, with the carrying capacities of each group determined as ratios of the total deme capacity $K_{\text{deme}}$ given above. We fix the intrinsic logistic growth rate $r$ = 1.3 per

generation, estimated from data on human population growth over the Holocene ~10 kya to pre-Industrial times (US Census Bureau: www.census.gov). The $F_D$ population size is also affected by selection in the presence of the LP allele (see below).

A unidirectional, and potentially long-distance, migration process is modelled independently for each cultural group. Each group *i* has a parameterized mobility factor $M_i$ (measured in demes) which allows us to model the likely differences between the migration patterns of the three groups. In each deme for cultural group *i* the migration destination deme is determined by a Gaussian random-walk process, which takes into account the mobility of the group and the relative mobility of the source deme by having standard deviation equal to the product $M_i M_{curr}$. This migration process allows individuals to make short sea-crossings, but if the chosen destination *is* a sea deme then migrants are instead placed in the nearest coastal deme. Alongside the larger coastal mobility factor $M_{curr}$, this process has the effect of increasing the rate of spread of farming along coastal routes, consistent with archaeological evidence (see for example Clark 1965).

A parameterized proportion of the individuals in each cultural group $P_{mig}$ are *available* to migrate. The actual number that migrate from the current to the destination deme is determined by the outward 'pressure' (i.e. how full the deme is), and the 'attractiveness' of the destination deme (i.e. the amount of available space). This number of migrants is calculated as follows:

$$N_{mig} = \frac{1}{2}\left(\frac{K_{dest} - N_{dest}}{K_{dest}} + \frac{N_{curr}}{K_{curr}}\right) P_{mig} N_{curr}$$

where $K_i$ and $N_i$ are the cultural group carrying capacities and population sizes respectively.

### 2.4.3 Genetic processes

We independently modelled two unlinked loci; one for the LP trait and one for a 'genetic background allele' – called GB hereafter – that was used to estimate the expected contribution of the initial LP-dairying gene-culture coevolutionary group to the modern European gene pool. The frequencies of the LP and GB alleles were initially set at zero everywhere, and we modelled the following genetic processes only after the LP allele arises in the model.

In each simulation we chose a random deme as the location for the start of LP / dairying gene-culture coevolution. When the size of the $F_D$ cultural group in this chosen deme reaches at least 20 the LP allele is initialized at a frequency of 0.1 (i.e. ~4 LP alleles). While we recognise that novel mutations *must* by definition arise at a frequency of $1/2N$, preliminary simulations initialized with just one allele copy very often led to extinction of the LP allele unless selection was very strong. Adopting this approach allowed us to seed the LP allele near the expanding population wavefront in a computationally efficient way. It is important to reiterate that this is *not* where the mutation giving rise to the LP allele *first* occurred, but rather is the location where the LP allele was first subject to a gene-culture coevolutionary forces in the presence of a dairying culture. In the same generation in which the LP allele is seeded, we also seed the unlinked GB 'allele' by assigning a frequency of 1 to all individuals in all cultural groups in the start location deme and its 8 neighbours.

We model four population genetic processes: (1) *inter*demic bidirectional gene flow; (2) *intra*demic bidirectional gene flow; (3) drift; and (4) selection acting solely on LP allele-carrying $F_D$ individuals. We assume Hardy Weinberg equilibrium in each cultural group in each deme. The first two processes operate on *both* alleles, but we do not model drift or selection for the GB allele in order for it to mimic the evolutionary history of rest of the genome.

(1) Interdemic gene flow models the exchange of individuals between the three cultural groups within a deme. A parameterized proportion of individuals in each cultural group $P_c$ are *available* to change group, and the actual number exchanged between group *i* and *j* is dependent on the relative population sizes $N_i$ and $N_j$ of each group. This is calculated as:

$$intra_{i \leftrightarrow j} = \frac{N_j}{N_i + N_j} P_c N_i$$

(2) Interdemic bidirectional gene flow models the exchange of individuals within a cultural group but between neighbouring demes. Again, a parameterized proportion of individuals in each cultural group $P_d$ are *available* to change deme. The actual number exchanged between the focal deme *i* and a single randomly chosen neighbouring deme *j* is now dependent on the relative population sizes $N_i$ and $N_j$ of the cultural group in each deme. This is similarly calculated as:

$$inter_{i \leftrightarrow j} = \frac{N_j}{N_i + N_j} P_d N_i$$

(3) Drift is modelled by simple generational binomial sampling of LP allele frequencies in each cultural group in each deme (following the emergence of the LP allele, which is explained below).

(4) Finally, selection is modelled *only in the dairying farmer* $F_D$ cultural group; this has two effects. The first, with parameterized selection strength $s_{F_D}$, is to increase the frequency of the LP allele $p$ in the group as follows (Maynard Smith 1998):

$$p' = \frac{p^2(1 + s_{F_D}) + pq(1 + s_{F_D})}{1 + s_{F_D}(p^2 + 2pq)}$$

where $p$' is the new LP allele frequency, $p$ is the previous frequency and $q = 1 - p$ (i.e. the previous frequency of non-LP individuals). But selection also has the demographic effect of increasing the population size of dairying farmers $F_D$ in each deme given the presence of the LP-allele. This increase, over and above the logistic growth described above, is calculated as follows (Maynard Smith 1998):

$$N' = N(1 + s_{F_D}(p^2 + 2pq))$$

where $N$' is the new population size of $F_D$ in the deme.

### 2.4.4 Cultural processes

The process of cultural diffusion models the spread of cultural practices by allowing individuals within a deme to change cultural group. A parameterized proportion of individuals in each cultural group $P_{\text{diff}}$ are *available* to change group. The number that actually convert from group *i* to group *j* is dependent on the weighted sum of the number of individuals in group *j* (as a proportion of carrying capacity) in the focal deme (0) and its mean value in the 8 neighbouring demes (1 to 8). This value $N_{\text{i,0}\rightarrow\text{j,0}}$ is calculated as:

$$N_{\text{i,0}\rightarrow\text{j,0}} = \left( b\frac{N_{\text{j,0}}}{K_{\text{j,0}}} + (1-b)\frac{1}{8}\sum_{\text{k=1}}^{8}\frac{N_{\text{j,k}}}{K_{\text{j,k}}} \right) P_{\text{diff}} N_{\text{i,0}}$$

where *b* is a parameter that weights the relative effect due to the focal deme and the 8 neighbouring demes. This is fixed at 0.75 throughout, meaning that the cultural composition of the focal deme has 3 times the influence of the neighbouring demes.

### 2.4.5 Model parameters

To summarize, the model has 7 processes, which involve a total of 16 parameters. While we kept all the *processes* independent for each cultural group, for ease of analysis we reduced this large number of parameters by assuming that the proportion available in each group to migrate or undergo gene flow or cultural diffusion could each be described by a common parameter. It should be noted that while the migration proportion parameter $P_{\text{mig}}$ is common to all three groups, the

mobility factor of each group is parameterized separately. The resulting 8 parameters of interest are given in table 2.1.

We have very little information from the archaeological record regarding the values most of these 8 parameters may take, so we constrained each parameter to a 'prior' range and in each simulation randomly assigned values from this uniform prior. This allowed us to utilize an approximate Bayesian approach to estimate the values of these parameters given the data (this is dealt with in greater detail in the following section). Prior ranges for each parameter are given in table 2.1. While not strictly modelled parameters, we also recorded, and made inferences on, the starting location and time ($T_{LP}$) for gene-culture coevolution of the LP allele, as well as the contribution of this initial group to the modern European gene pool as tracked by the GB 'allele'. This contribution $F_{GB}$ was calculated in the final generation of each simulation as follows:

$$F_{GB} = \frac{\sum_{i} \sum_{j} p_{GB_{i,j}} N_{i,j}}{\sum_{i} N_{i}}$$

where *i* iterates over the whole set of demes in the model, and *j* iterates over the set of cultural groups {$F_D$, $F_{ND}$, $HG$}. $N_{i,j}$ and $p_{GB_{i,j}}$ are the population size and GB allele frequency in cultural group *j* in deme *i* respectively, while $N_i$ is the total population size of deme *i*.

### 2.4.6 Simulation process

Each simulation was initialized at 9,000 ya with $F_D$ and $F_{ND}$ populating Anatolia and *HG* across the whole of western Eurasia, and we simulated forwards to the present day, i.e. for 360 generations assuming a 25 year generation time (Thomas *et al.* 2006, Tremblay and Vezina 2000). We performed 200k simulations in total. In the final generation of each simulation we recorded the LP allele frequencies and farming arrival times at 12 different locations around western Eurasia. The time to the arrival of farming was defined as the time / generation at which either $F_D$ or $F_{ND}$ reached 1% of their carrying capacity within a sample location deme. This simulated data was then compared to observed -13,910*T allele frequencies and carbon-14 based estimates of the arrival of farming at the same 12 locations (see table 2.2, Bersaglieri *et al.* 2004, Mulcare 2006, Pinhasi *et al.* 2005). More detail on the observed data is provided in section 2.6.

This is a relatively complex model compared to previous human demographic / evolutionary models (Barbujani *et al.* 1995, Ray *et al.* 2003, Excoffier 2004, Currat and Excoffier 2005). In order to use this genetic and archaeological data in an integrated way to estimate values for the model parameters we used a statistical approach called approximate Bayesian computation (ABC). This is a rapidly developing family of techniques that allows for the robust analysis of complex, parameter-heavy models. The following section provides a mathematical overview of some inference methods commonly used in population genetic modelling, with a firm emphasis on ABC

**Table 2.1:** The model parameters and their prior ranges. The last two are pseudo-parameters that arise in the model, with pseudo-priors that span the entire possible range.

| parameter | | prior range |
|---|---|---|
| migration proportion | $P_{\text{mig}}$ | [0, 0.2] |
| *HG* mobility | $M_{\text{HG}}$ | [0, 3] demes |
| $F_{\text{ND}}$ mobility | $M_{F_{\text{ND}}}$ | [0, 3] demes |
| $F_{\text{D}}$ mobility | $M_{F_{\text{D}}}$ | [0, 3] demes |
| interdemic gene flow | $P_{\text{d}}$ | [0, 0.2] |
| intrademic gene flow | $P_{\text{c}}$ | [0, 0.2] |
| cultural diffusion | $P_{\text{diff}}$ | [0, 0.2] |
| selective advantage to $F_{\text{D}}$ | $s_{F_{\text{D}}}$ | [0, 0.2] |
| LP gene-culture coevolution start time | $T_{\text{LP}}$ | [0, 9000] ya |
| contribution to modern European gene pool | $F_{\text{GB}}$ | [0, 100] % |

## .2.5 Statistical inference and approximate Bayesian computation (ABC)

### 2.5.1 Likelihood

The initial motivation of statistical inference is that we have observed data *D*, which we assume to have been generated under a model specified by a set of parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. We wish to make inferences about these parameters $\theta$,

as well as evaluating the ability of a model to explain the data in comparison to alternative models. The most statistically robust way to do this is to calculate the conditional probability $P(D|\theta)$ of obtaining the observed data under a specified model with parameter values $\theta$. When considered as a function of the parameters $\theta$ with $D$ fixed (i.e. observed) this is called the 'likelihood' $L(\theta; D)$, and is proportional to $P(D|\theta)$.

The likelihood function has long been of central importance in statistical inference as it captures all the available information on the parameters contained in the data, the so-called 'likelihood principle'. The best estimate of the parameters given the data, the *maximum likelihood estimator* (MLE) $\hat{\theta}$, can be found by maximizing this likelihood function. Competing models, or hypotheses, can also be evaluated using the *likelihood-ratio test*, which allows the rejection of a null model, say $H_0$, in favour of an alternative, say $H_1$. The null model $H_0$ can be rejected if the test statistic $L(\hat{\theta}_0)/L(\hat{\theta}_1)$, where $\hat{\theta}_i$ is the MLE for model $H_i$, is significant, i.e. the likelihood of the alternative $H_1$ is significantly larger. This test statistic is usually calculated in terms of *log*-likelihoods, i.e. $-2\log(L(\hat{\theta}_0)/L(\hat{\theta}_1))$, which under the null model is $\chi^2$-distributed.

### 2.5.2 Difficulties with full likelihood approaches

However, for anything more complex than the very simplest of models, direct calculation of the likelihood function is usually extremely difficult, if not impossible. In population genetics we are often interested in estimating parameters from increasingly complex demographic models, which are usually based in coalescent

theory (see Chapter 6). This means that calculation of the associated likelihood function involves high-dimensional summation or integration over all possible genealogical histories for the sample of observed data, which rapidly becomes intractable as sample size increases (see Felsenstein 1988).

### 2.5.3 Estimating the likelihood – MCMC and Importance Sampling

Two main methodological approaches to overcome this problem have been developed, both of which employ Monte Carlo simulation to *estimate* the likelihood function. Monte Carlo simulation (see Metropolis *et al.* 1953) is a technique that allows the approximate numerical calculation of high-dimensional integrals through repeated independent random sampling of the integrand, which in a population genetic context represents the space of all possible genealogies and parameter values. However, an overwhelming proportion of this often astronomically large parameter space is very unlikely, i.e. contributes only negligibly to the likelihood function. The two approaches I describe here provide a way of efficiently exploring just those parts of parameter space that have non-negligible likelihood.

The first approach couples Monte Carlo simulation with a Markov chain (MCMC) (see Metropolis *et al.* 1953, Hastings 1970). An MCMC analysis is initialized with parameter values $\theta_0$, and then a chain of *dependent* samples from parameter space $\theta_0,\theta_1,\theta_2,\ldots,\theta_n$ is created by moving from state $\theta_i$ to state $\theta_j$ subject to a transition probability. This probability is the product of the ratio of predefined 'proposal probabilities' of moving between $\theta_i$ and $\theta_j$ (which may be asymmetric, see Hastings

1970), and the likelihood ratio $P(D|\theta_j)/P(D|\theta_i)$, which is obtained by simulating under the model with parameter sets $\theta_i$ and $\theta_j$. This means, in generic terms, that such a Markov chain spends more of its time in the more likely parameter space, and its long-term equilibrium distribution approximates the full likelihood distribution. This method was first applied in a population genetic / coalescent context by Kuhner *et al.* (1995, 1998, see also Wilson and Balding 1998, Beaumont 1999, Beerli and Felsenstein 1999 and Nielsen and Wakeley 2001) to estimate the composite population mutation parameter $\theta = 4N_e\mu$ from DNA sequence data, where $N_e$ is the effective population size and $\mu$ the mutation rate.

The second approach, importance sampling (IS) (see Ripley 1987) allows, in generic terms, efficient estimation of an unknown target distribution by taking samples from a carefully transformed version of the target with frequencies in proportion to their weight or 'importance' in the original target. In our context, this method allows the creation of independent samples from the underlying likelihood function, meaning it can be much more efficient than the MCMC sampling process which has to have reached equilibrium – something that can take a long time and / or can be difficult to guarantee. This approach was pioneered in population genetics by Griffiths and Tavaré (1994a, 1994b, see also Stephens and Donnelly 2000).

Both of these methods provide an estimate of the likelihood function, so inference on the parameters can then be carried out either through maximum likelihood, or by embedding them in an explicitly Bayesian framework (see Wilson and Balding 1998 and Beaumont 1999).

### 2.5.4 The Bayesian framework

Under the Bayesian paradigm the parameters of a model are themselves assumed to have a probability distribution, to reflect the uncertainty in their true values. The distribution of the parameters *before* the data has been observed $P(\theta)$ is called the *prior* distribution, and allows the inclusion of previous knowledge about the parameters unrelated to the data (which may be objective or reflect the subjective beliefs of the researcher). After observing the data $D$, we can 'update' our prior knowledge about the parameters with information from the data. This conditional distribution of the parameter given the data $P(\theta \mid D)$ is called the *posterior* distribution, and is calculated using Bayes' theorem:

$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}$$

where $P(D \mid \theta)$ is the likelihood, and the denominator $P(D)$ is a normalizing constant (formally the 'prior probability of the data' $= \int P(D \mid \theta)P(\theta)d\theta$).

All inferences about the parameters $\theta$ are derived from this posterior distribution. The mode of the marginal posterior distribution for each parameter can be taken as a point estimate, and credible intervals (CIs) are easily constructed using the tail-quantiles of the distribution. An $x$% Bayesian CI can be simply interpreted as containing the true parameter with probability $x$%. This is in contrast to an $x$% frequentist confidence interval ($CI_f$), which gives the percentage of identically created $CI_f$ from repeated sampling that will contain the true parameter. For posterior

distributions that are not unimodal, *x*% highest posterior density (HPD) intervals are often given, which are simply the intervals containing *x*% of the highest posterior probability.

The major benefit of the Bayesian framework arises from its key difference with frequentist approaches, namely that it considers the unknown parameters as random variables with probability distributions. Rather than providing probability statements on the data as in frequentist approaches, Bayesian methods provide statements on the unknown parameters in light of the data already observed. The only substantive criticism of the Bayesian framework concerns the reliance on a prior distribution. Choosing non-informative priors that represent total ignorance about the parameter can be difficult or even lead to inconsistencies, and there is also controversy over how subjective choices of prior may impact on scientific inference. However, these problems are largely technical issues and do not detract significantly from the fact that through the incorporation of a prior the Bayesian framework provides a natural way of utilizing *all* the information available. Now that some of the previous computational issues (as briefly outlined above) that limited their applicability are beginning to be overcome, Bayesian techniques have become increasingly prevalent and important in genetic analysis (see Beaumont and Rannala 2004 for a recent review).

### 2.5.5 Approximate Bayesian computation (ABC)

#### 2.5.5.1 ABC outline

However, whether we adopt a Bayesian approach or not, the central issue is that the necessary estimation of likelihood function can be difficult and time-consuming, especially as models become more complex. A recent, and ongoing, development known as approximate Bayesian computation (ABC) neatly circumvents the explicit consideration of likelihood functions. In brief, ABC produces approximate posterior distributions for the parameters of a model by repeatedly simulating with parameter values chosen randomly from prior distributions and accepting only those values that result in a 'good fit' between the simulated and observed data. This approach was first outlined by Tavaré *et al.* (1997), and important subsequent extensions to this scheme (Fu and Li 1997, Weiss and von Haeseler 1998, Pritchard *et al.* 1999, Beaumont *et al.* 2002) have allowed its application to a wide variety of inference problems both in population genetics (see for example Estoup *et al.* 2004, Fagundes *et al.* 2007, Ray *et al.* 2010) and elsewhere (see for example Toni and Stumpf 2010).

#### 2.5.5.2 Development of ABC

The origin of ABC lies in a 'rejection' algorithm due to Tavaré *et al.* (1997) that allows the creation of samples from the posterior distribution. In order to introduce the full ABC algorithm, consider the following sampling scheme:

(1) Draw parameter values $\theta^{(i)}$ from their prior distribution $P(\theta)$, where (i) is the sample index.

(2) Accept parameters $\theta^{(i)}$ with probability proportional to $P(D|\theta)$, i.e. the likelihood. Return to step (1)

The accepted parameter values are distributed exactly as the posterior $P(\theta|D)$. This is simply a restatement of Bayes' theorem in the form of an iterative algorithm; we update our prior knowledge about the parameters using the likelihood. However, in order to avoid the calculation of the likelihood function $P(D|\theta)$, Tavaré *et al.* (1997) demonstrated that the target posterior $P(\theta|D)$ can be obtained by generating simulated data under a model (with parameters sampled from priors) and accepting just those parameter sets for which the simulated data matches the observed data. This was further generalized by Fu and Li (1997), whose use of Monte Carlo simulation allowed this approach to be extended to far more complex demographic models. The algorithm is as follows:

(1) Draw parameter values $\theta^{(i)}$ from their prior distribution $P(\theta)$.

(2) Generate simulated data $D^{(i)}$ under a stochastic model with parameters $\theta^{(i)}$.

(3) Accept parameters $\theta^{(i)}$ if the simulated data equals the observed data, i.e. $D^{(i)} = D$. Return to step (1).

This produces independent samples from the target posterior $P(\theta|D)$, so repeated iteration will result in a distribution of accepted parameter values that converges exactly to $P(\theta|D)$. Both Tavaré *et al.* (1997) and Fu and Li (1997) suggested replacing the full data $D$ with summary statistic $S$. These summary statistics should ideally be 'sufficient' for the data, meaning that they capture all relevant information on the parameters in the data, i.e. $P(\theta|D) = P(\theta|S)$. The choice of sufficient statistics can be problematic, as no formal method exists; this will discussed in more detail below. Weiss and von Haeseler (1998) extended the Monte Carlo simulation approach to incorporate multiple statistics and more than a single parameter. They also relaxed the stringent acceptance criterion of an exact match between the statistics calculated on the simulated $s^{(i)}$ and observed data $s$, by retaining parameter sets when the 'distance' between the two vectors of statistics $\rho(s^{(i)}, s)$ was less than a small arbitrary tolerance level $\delta$.

#### 2.5.5.3 Rejection ABC

The full ABC 'rejection' algorithm was clearly set out by Pritchard *et al.* (1999) as follows:

(1) Choose summary statistics $S$ and calculate these on the observed data $s$.

(2) Choose a tolerance $\delta$.

(3) Draw parameter values $\theta^{(i)}$ from their prior distribution $P(\theta)$.

(4) Generate simulated data $D^{(i)}$ with parameters $\theta^{(i)}$.

(5) Calculate summary statistics on the simulated data $s^{(i)}$.

(6) Calculate distance measure $\rho(s^{(i)}, s)$.

(7) Accept parameters $\theta^{(i)}$ if $\rho(s^{(i)}, s) \leq \delta$. Return to step (1).

The accepted simulations are independent observations from $P(\theta \mid \rho(s^{(i)}, s) \leq \delta)$. As $\delta \to \infty$ this process simply recaptures the prior distribution, but as $\delta \to 0$ it asymptotically approaches $P(\theta \mid S)$, which is equal to our target posterior $P(\theta \mid D)$ given that the statistics are sufficient. However, this rejection algorithm is not particularly efficient. The tolerance $\delta$ has to be small enough for the accepted simulations to approximate the posterior (rather than being dominated by the prior), meaning that the overwhelming majority of simulations, typically many millions, are rejected for just a few thousand acceptances. So the choice of $\delta$ is a trade-off between accuracy of the approximation to the posterior and computational efficiency.

#### 2.5.5.4 Regression adjusted ABC – Beaumont *et al.* (2002)

A way to improve computational efficiency is to allow larger tolerance $\delta$, i.e. accept more (and 'worse') simulations, but then adjust the retained parameters to explicitly take account of the actual distance $\rho(s^{(i)}, s)$ (i.e. 'quality of fit') for each set rather than using it simply as a Boolean indicator for acceptance / rejection. This approach was first proposed by Beaumont *et al.* (2002). They suggested – as algorithm steps (8) and (9) – that the retained parameter sets $\theta^{(i)}$ be smoothly weighted according to $\rho(s^{(i)}, s)$ and then adjusted with a multiple local-linear

regression which corrects for the discrepancy between $s^{(i)}$ and $s$. This is achieved by using an Epanechnikov kernel weighting $K_\delta$ in the regression:

$$K_\delta(\rho) = \begin{cases} c\delta^{-1}(1-(\rho/\delta)^2) & \rho \le \delta \\ 0 & \rho > \delta \end{cases}$$

where $c$ is a normalizing constant and $\rho$ is shorthand for the *magnitude* of the distance $\rho(s^{(i)}, s)$. While other kernels can be used, the steep decline to zero of this function near $\rho = \delta$ ensures that very few accepted parameters are given small non-zero values, thus aiding computational efficiency (see Beaumont *et al.* 2002). They also proposed for the distance measure $\rho$ the Euclidean norm between the two (variance normalized) vectors of statistics $s^{(i)}$ and $s$, i.e. $||s^{(i)} - s|| = \sqrt{\sum_j (s_j^{(i)} - s_j)^2}$, where $j$ is the index of each statistic within the set $S$. So the acceptance region is the $n$-dimensional sphere with radius $\delta$ (the 'δ-ball') centered at $s$. Also, in practice a tolerance proportion $F_\delta$ is often pre-defined, which gives the quantile of simulations to accept and is used to implicitly calculate the absolute tolerance value δ.

Beaumont *et al.* (2002) demonstrated that including these additional post-rejection adjustment steps greatly reduced the relative mean square error (rMSE) of the parameter estimates compared to the previous rejection method. They showed that the quality of estimation was largely insensitive to increasing tolerance proportion $F_\delta$, while for the rejection method the rMSE increased rapidly as more simulations were accepted. Even in comparison to the full-likelihood MCMC method of Wilson and Balding (1998), regression-adjusted ABC performed well, but with

orders of magnitude fewer simulations. They concluded that ABC is a valid and efficient alternative to the previous computationally intensive MCMC approaches, and it has since been successfully used to analyze relatively complex demographic models (see for example Estoup *et al.* 2004, Hamilton *et al.* 2005, Fagundes *et al.* 2007, Neuenschwander *et al.* 2008). It should be noted that the ABC algorithm I have implemented in this chapter and Chapter 6; section 6.5 is based primarily on that of Beaumont *et al.* (2002) as just described.

One potential problem with this regression adjustment is that the resulting posterior distribution can be non-zero for values of the parameter for which the prior is zero, i.e. the posterior *can* be projected outside the range of the prior. Recently Leuenberger and Wegmann (2010) have suggested an alternative regression step that avoids this problem, which involves fitting a general linear model (GLM) to adjust the parameters. This adjustment, called ABC-GLM, perhaps more naturally considers the summary statistics as response variables rather than explanatory as in Beaumont *et al.* (2002), and provides a marginal improvement in posterior estimation (though this improvement is larger if the priors are more complex). Blum and François (2009) have also proposed extending the regression step of Beaumont *et al.* (2002) to consider more complicated non-linear models, in order to account for likely heteroscedasticity in the regression error terms.

#### 2.5.5.6 Improving the efficiency of ABC

Even though these post-rejection adjustment methods allow an increase in tolerance δ, a large number of simulations still have to be carried out, only to be rejected. Other recent extensions have attempted to improve on the inefficiency inherent in random sampling of the prior distribution in step (1) of the algorithm by using an MCMC approach. In one such method, ABC-MCMC (Marjoram *et al.* 2003, see also Wegmann *et al.* 2009), the Markov chain explores parameter space more efficiently than rejection-based methods and converges to the target approximate posterior distribution. Further improvements on this MCMC-ABC scheme based on sequential (Sisson *et al.* 2007, Toni *et al.* 2009) or population (Beaumont *et al.* 2009) Monte Carlo sampling have been suggested, though a consensus is yet to be reached.

#### 2.5.5.7 Statistic choice

Perhaps more important than computational inefficiency, one outstanding problem with the ABC approach (as noted above) concerns the choice of summary statistics *S*. The choice of sufficient statistics is difficult to guarantee, and in their absence the degree to which the ABC-generated distribution $P(\theta \mid \rho(s^{(i)}, s) \leq \delta)$ approximates the true posterior $P(\theta \mid D)$ is unknown. In principal one solution might seem to be to take a large, even exhaustive, set of summary statistics *S* in the hope that *S* will 'cumulatively approach' sufficiency. But in practice, the 'curse of dimensionality' means that increasing the number of statistics will generally increase $||s^{(i)} - s||$ and thus reduce the number of acceptances for a given tolerance δ, i.e.

reduce the efficiency (Joyce and Marjoram 2008). So including statistics that are highly correlated incurs a cost for very little gain in information on the data, and the inclusion of uninformative statistics also serves to increase the noise in the calculation of $||s^{(i)} - s||$. Joyce and Marjoram (2008) propose an algorithm to provide 'approximate sufficiency', by sequentially adding statistics to a set and retaining them if they contribute significantly to the quality of inference, as measured by the odds-ratio of this new augmented set against the pre-existing set. They showed that this approach provided an improvement in parameter estimation, by reducing the mean square error (MSE).

Wegmann *et al.* (2009) have suggested using partial least squares (PLS) regression to transform the set of chosen summary statistics prior to any analysis (see Ray *et al.* 2010 for an example). Similar to principal components analysis (PCA), this reduces the dimension of a large and potentially correlated set of statistics by creating a smaller orthogonal set of components that still explain most of the variation in the parameters of the model. This also reduces the noise in the distances $||s^{(i)} - s||$ as uninformative statistics are removed i.e. they have negligible or zero weight in each component.

#### 2.5.5.8 ABC Summary

ABC, and the unresolved choice of statistics problem in particular, is currently a very active area of research in statistical genetics, and statistics in general. While I have provided a brief overview of some of the latest developments, it is important to reiterate that the model analysis presented in the following section, and Chapter 6;

section 6.5, is based on the ABC algorithm given above, as canonically described by Beaumont *et al.* (2002).

## 2.6 Simulation model ABC analysis and results

### 2.6.1 ABC analysis

In analyzing this model we adopted the approach, also employed by Sousa and colleagues (Sousa *et al.* 2009), of using the raw -13,910*T allele frequencies directly as summary statistics, though we further included statistics on the arrival time of farming in Europe. Our full set of summary statistics *S* comprised: the -13,910*T allele frequencies at the 12 different sample locations around western Eurasia (see table 2.2); the time in generations to arrival of farming at 11 of the same locations (excluding Anatolia as it served as the origin of the spread of farming into Europe, i.e. generation 0) (see table 2.2); and two Spearman's rank-order correlation coefficients, calculated separately for the 12 -13,910*T allele frequencies and the 11 times to arrival of farming. This gave a total of 25 summary statistics.

The farming arrival time statistics were estimated using calibrated carbon-14 data from Pinhasi *et al.* (2005). This estimation procedure involved the following steps: (1) the mean nearest-neighbour distance between the sampling locations $\bar{r}_E$ was calculated (557.13 km); (2) A Gaussian sampling region was constructed around each of the 11 sampling locations, with standard deviation of $\bar{r}_E/1.96$ so that 95% of each sampling region would fall within $\bar{r}_E$; (3) A weighted average of all calibrated carbon-14 earliest farming arrival dates from Pinhasi *et al*. (2005) within 3 standard

deviations of each sampling location was calculated, weighted by the distance from the sampling location and the Gaussian density. Assuming a generation time of 25 years (Thomas *et al.* 2006, Tremblay and Vezina 2000), these estimated arrival dates were scaled to generations from the start of the simulation, which was set at 9,000 years or 360 generations ago (see table 2.2).

**Table 2.2:** Details of the 12 sample locations. References for -13,910*T allele frequencies: M = Mulcare (2006), B = Bersaglieri *et al.* (2004). Times of the arrival of farming in years before present were estimated from Pinhasi *et al.* 2005 (see text for details). Generations in the model starting from the origin 9,000 years or 360 generations ago (assuming a 25 year generation time) are given in brackets.

| **sample location** | **latitude, longitude** | **-13,910*T allele frequency (ref)** | **farming arrival time (ya) (generation)** |
|---|---|---|---|
| Anatolia | 38.00, 30.00 | 0.031 (M) | 9000 (0) |
| Greece | 37.98, 23.73 | 0.134 (M) | 8389 (24) |
| Tuscany | 43.77, 11.25 | 0.063 (B) | 7112 (76) |
| Sardinia | 39.00, 9.00 | 0.071 (B) | 6968 (81) |
| North Italy | 45.68, 9.72 | 0.357 (B) | 6911 (84) |
| Scandinavia | 59.33, 18.05 | 0.815 (B) | 6197 (112) |
| Germany | 53.55, 10.00 | 0.556 (B) | 6434 (103) |
| France | 48.87, 2.33 | 0.431 (B) | 6197 (112) |
| French Basque | 43.00, -1.00 | 0.667 (B) | 6037 (119) |
| Southern UK | 51.50, -0.12 | 0.734 (M) | 5905 (124) |
| Orkney | 58.95, -3.30 | 0.688 (B) | 5306 (148) |
| Ireland | 54.37, -7.63 | 0.954 (M) | 5260 (150) |

The parameters of interest $\theta$ are*:* migration proportion $P_{mig}$; mobility of each of the three cultural groups $M_i$; interdemic gene flow proportion $P_d$; intrademic gene flow proportion $P_c$; cultural diffusion proportion $P_{diff}$; selective advantage of LP within the $F_D$ group $s_{F_D}$; generation at which gene-culture coevolution of the LP allele starts (later reconverted to time in years) $T_{LP}$; east-west and north-south coordinates of this start location; and the contribution of individuals in this location to the modern European gene pool $F_{GB}$. The uniform prior distributions for each parameter are given in table 2.1.

The full ABC algorithm I implemented, in the Python programming language (van Rossum *et al.* URL: http://www.python.org/), was as follows: (1) choose the summary statistics $S$ as discussed above and calculate their values $s$ for the observed data (see table 2.1); (2) pre-define a proportion of the best-fitting simulations $F_\delta$ to accept, and from this calculate an implicit tolerance level δ; (3) draw a parameter set $\theta^{(i)}$ from the prior distributions of $\theta$, where (i) is an index; (4) simulate forward under the model using parameter set $\theta^{(i)}$; (5) in the final / 360$^{th}$ generation calculate the summary statistics $s^{(i)}$ for the simulated data; (6) accept the simulated parameter set $\theta^{(i)}$ if $||s^{(i)} - s|| \leq \delta$, where $||.||$ is the Euclidean norm between the two vectors; (7) repeat steps 3 to 6 until a sufficient number of parameter sets have been retained i.e. $F_\delta \times 200{,}000$; (8) perform a local-linear standard multiple regression to adjust the $\theta^{(i)}$, with each $\theta^{(i)}$ weighted according to the size of $||s^{(i)} - s||$ using the Epanechnikov kernel function $K_\delta$. The resulting fitted parameter sets $\theta^{(i)*}$ constitute independent random samples from the approximate joint posterior distribution $P(\theta | S = s)$, from which marginal distributions can be obtained.

All retained parameters – except for the two coordinate values and the generation at which the coevolutionary process starts – were log transformed prior to the regression step, and subsequently back-transformed to produce the fitted parameter sets $\theta^{(i)*}$, as suggested by Beaumont *et al*. (2002).

## 2.6.2 ABC results

The tolerance level $F_\delta$ was set at 0.5%, meaning we accepted the 1,000 best-fitting simulations, though the main results were robust to different $F_\delta$ values. Marginal posterior parameter estimates are given in figure 2.2, and table 2.3 gives posterior modes and 95% credible intervals (CIs). The model and the observed data appear informative on a number of parameters, as outlined below.

### 2.6.2.1 Posterior distributions for parameters

For the selective advantage of the LP allele within dairying farmers $s_{F_D}$ the posterior 95% CI was considerably narrower (0.0518 to 0.159, mode = 0.0953) than its prior distribution (0 to 0.2), and in good agreement with previous genetic estimates for Europeans (0.014 to 0.19) (Bersaglieri *et al.* 2004). The 95% CI for the intrademic gene flow parameter $P_c$ is at the lower end (0.00206 to 0.0867, mode = 0.0153) of its prior distribution (0 to 0.2), which suggests a low level of inter-cultural gene flow between farming groups and hunter-gather groups.

A further notable result is that the inferred mobility of dairying farmers $M_{F_D}$ (mode = 3.12) is significantly higher than that of non-dairying farmers $M_{F_{ND}}$ (mode = 0.733). However, the observed data appears to contain little information on the mobility of the hunter-gatherers $M_{HG}$, as its posterior distribution is broader. This is not surprising, as the mobility of the relatively small *HG* population is unlikely to significantly affect the modern distribution of the LP allele, nor the times to arrival of farming. It should be noted that some posterior CI intervals were projected outside of the prior range, as a consequence of the regression adjustment as mentioned in the previous section.

#### 2.6.2.2 The origin of LP / dairying coevolution

The most probable location for the start of LP / dairying coevolution was estimated to lie in the region between the central Balkans and central Europe (figure 2.3). We also estimated a narrow range of times $T_{LP}$ at which this process began (figure 2.4a), with a marginal posterior mode of 7,441 ya (95% CI: 6,256 to 8,683 ya). We have made available three animations created from the best-fitting simulations that show plausible scenarios for the spread of the LP allele across Europe through time (Itan *et al.* 2009; videos S1 to S3).

**Table 2.3:** Marginal posterior estimates of demographic and evolutionary parameters; mode and 95% credibility interval (CI). Posterior distributions were by estimated by ABC employing regression adjustment and weighting of simulations accepted at tolerance level $F_{\delta}$ = 0.5%.

| **parameter** | | **prior range** | **posterior mode** | **posterior 95% CI** |
|---|---|---|---|---|
| migration proportion | $P_{\text{mig}}$ | [0, 0.2] | 0.129 | [0.0575, 0.251] |
| *HG* mobility | $M_{\text{HG}}$ | [0, 3] | 1.16 | [0.333, 2.17] |
| $F_{\text{ND}}$ mobility | $M_{\text{F}_{\text{ND}}}$ | [0, 3] | 0.733 | [0.311, 1.18] |
| $F_{\text{D}}$ mobility | $M_{\text{F}_{\text{D}}}$ | [0, 3] | 3.12 | [2.15, 3.93] |
| interdemic gene flow | $P_{\text{d}}$ | [0, 0.2] | 0.0440 | [0.00716, 0.171] |
| intrademic gene flow | $P_{\text{c}}$ | [0, 0.2] | 0.0153 | [0.00206, 0.0867] |
| cultural diffusion | $P_{\text{diff}}$ | [0, 0.2] | 0.0136 | [0.00113, 0.0847] |
| selective advantage to $F_{\text{D}}$ | $s_{\text{F}_{\text{D}}}$ | [0, 0.2] | 0.0953 | [0.0518, 0.159] |
| LP gene-culture coevolution start time (ya) | $T_{\text{LP}}$ | [0, 9000] | 7441 | [6256, 8683] |
| contribution to modern European gene pool (%) | $F_{\text{GB}}$ | [0, 100] | 7.47 | [2.83, 27.4] |

**Figure 2.2:** Approximate marginal posterior density estimates of the 8 demographic and evolutionary parameters. The boundaries of the 95% CIs are given by shading.

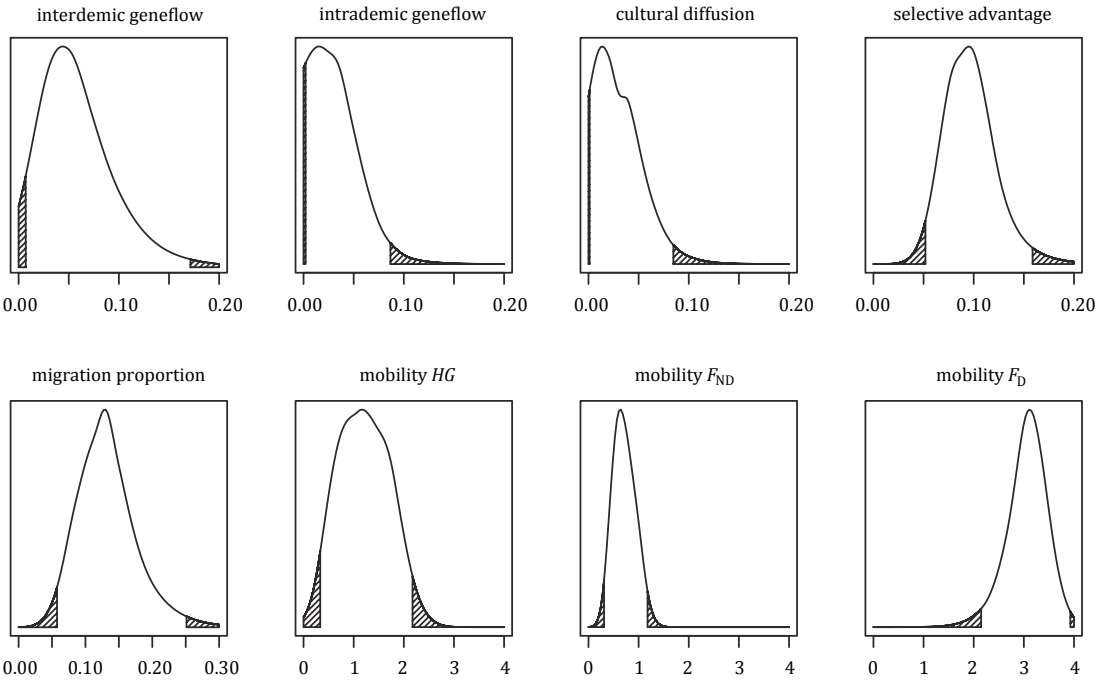


**Figure 2.3:** Approximate posterior density of location of origin for LP / dairying gene-culture coevolution. Points represent regression-adjusted latitude and longitude coordinates, and shading was added using kernel density smoothing.

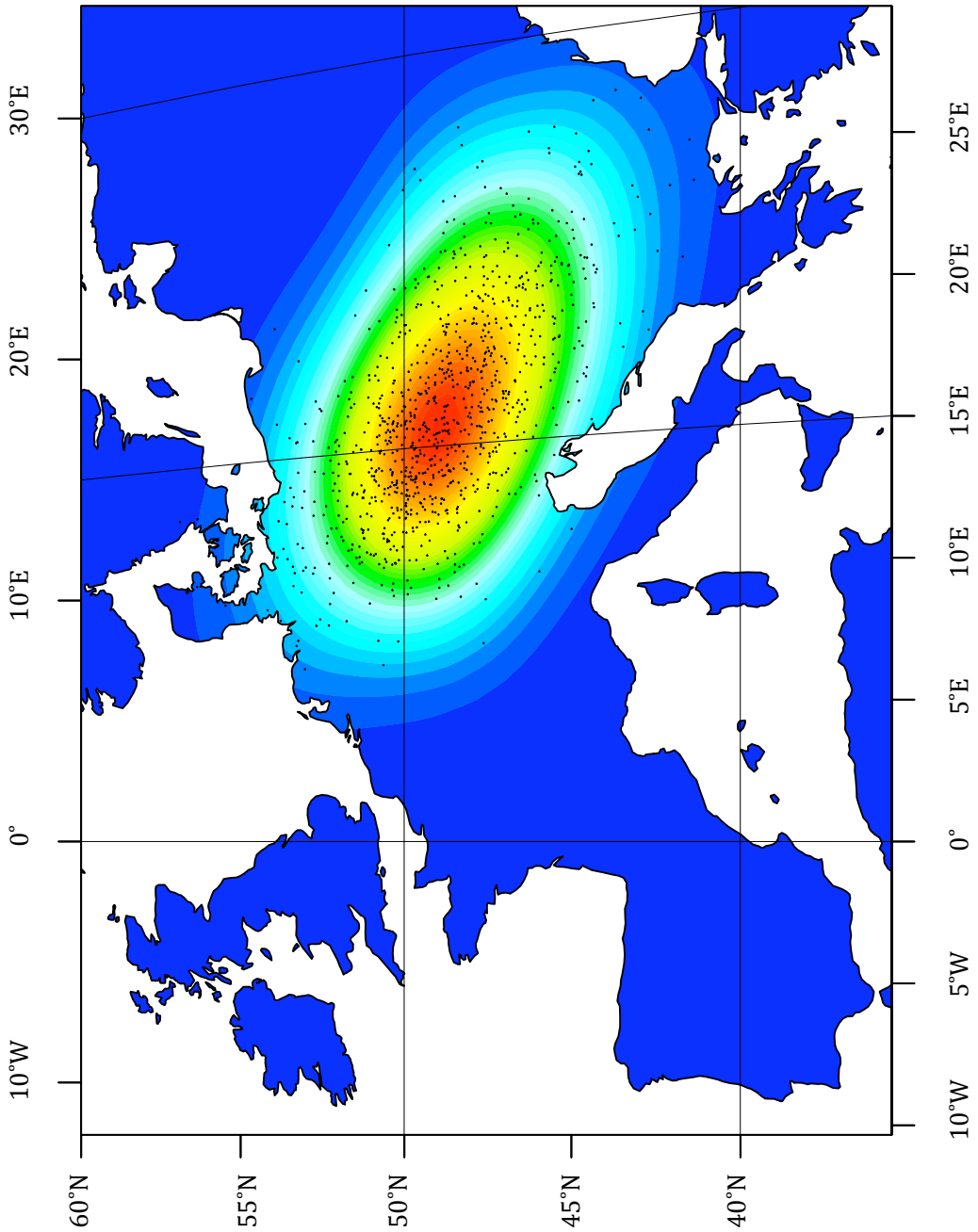
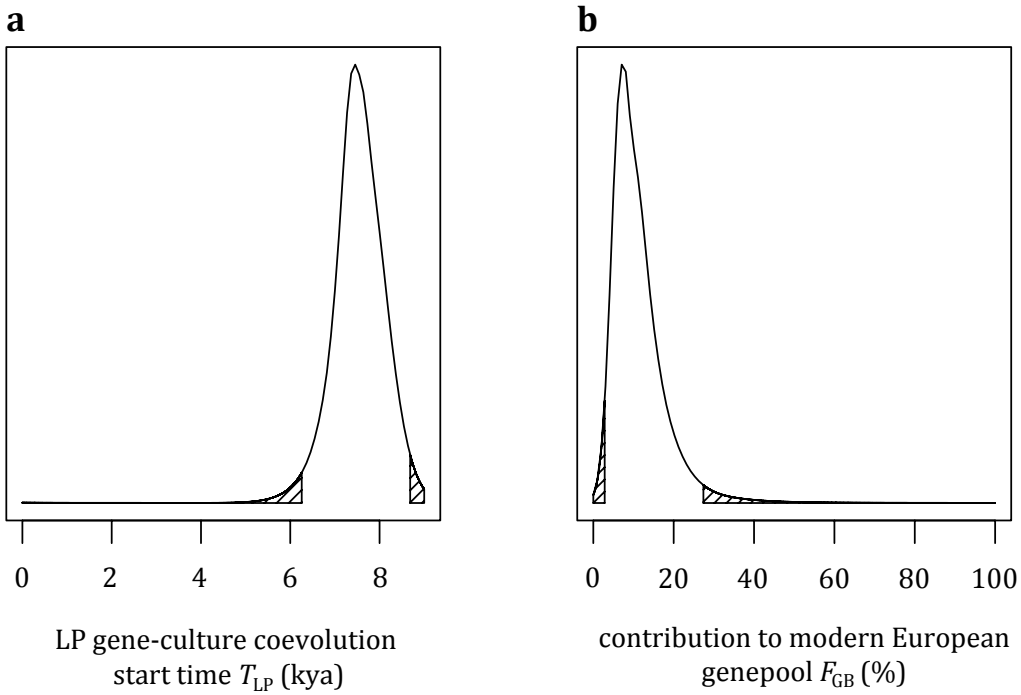


**Figure 2.4: a** The date of origin for LP / dairying coevolution (kya). **b** The percentage contribution of individuals living in the location of origin for LP / dairying coevolution to the modern European gene pool. The boundaries of the 95% CIs are given by shading.




### 2.6.2.3 The genetic contribution to the modern European gene pool

The genetic contribution to the modern European gene pool of the initial LP / dairying population, as estimated by the pseudo-parameter $F_{GB}$, also had a narrow marginal posterior (figure 2.4b); with mode 7.47% (95% CI: 2.83% to 27.4%). The demographic expansion process we simulated meant that this value was largely dependent on the LP / dairying coevolution start location. The closer this start

location was to the origin of the farming expansion in the southwest of the continent the larger the expected contribution to European ancestry would be.

In order to investigate whether the coevolution of LP and dairying increased the relative ancestral contribution of the initial group, we performed two further sets of 5,000 simulations each; one in which we allowed selection on LP in dairying farmers and one in which we did not. We initialized each simulation in each set with identical parameter values, which were sampled at random from the marginal posteriors already estimated, except that in the second we fixed the selective advantage to zero. Comparing the genetic contributions made by the initial LP / dairying population with and without selection we were surprised to find very little difference in the distributions between the models (see Itan *et al*. 2009; figure 5).

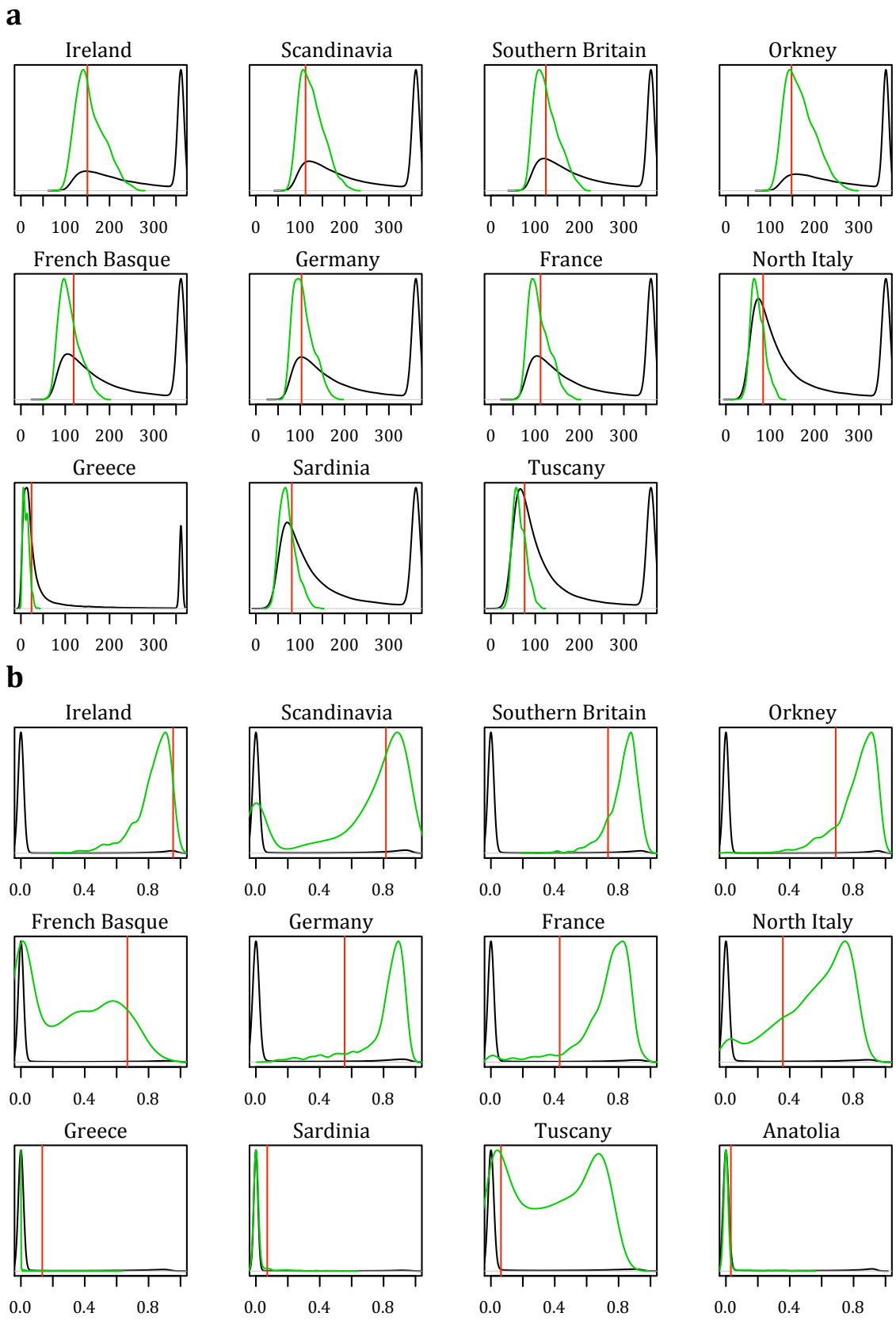
#### 2.6.2.4 Performance of the model

To evaluate the performance of our simulation model in explaining the observed data, we compared the distribution of each statistic generated by the simulations accepted for ABC analysis to the observed statistic at each sample location. For each location and for both farming arrival time and 13,910*T / LP allele frequency we plotted: (1) the observed value (table 2.2); (2) the distribution of the statistic obtained from all simulations in which the LP allele arose but did not go extinct; and (3) the distribution obtained from the simulations accepted at 0.5% tolerance for the ABC analysis. Although it will necessarily be the case that the 0.5% closest points will be nearer to the observed summary statistics than those simulated from the prior, it is still possible that the observed value will be an outlier from the

distribution of simulated points, which would indicate that the model is a poor fit. However, for the farming arrival times, simulations accepted for ABC analysis at the 0.5% tolerance level generate narrow ranges of expected values, in very good accordance with the observed values (figure 2.5a). This indicates that with our ABC-estimated parameter values, our model explains the farming arrival dates very well.

For the 13,910*T / LP allele frequencies, all observed values fall within the 95% equal tail probability interval of expected values from the simulations accepted at the 0.5% tolerance level (figure 2.5b). While a number of the expected modes are offset from the observed values, it is notable that for most of the northern European locations the expected mode is greater than the observed, while the opposite is the case for the southern locations (figure 2.5b).

**Figure 2.5 (next page):** Performance of model in explaining observed data on: **a** the estimated time of arrival of farming at 11 locations throughout Europe; **b** -13,910*T / LP allele frequency at 12 locations throughout Europe. The observed values are indicated by vertical red lines. The distributions of expected values from all simulations in which the LP allele arose and did not go extinct are indicated by black lines. The distributions of expected values from simulations accepted at the 0.5% tolerance level in ABC analysis are indicated by green lines.


a
Ireland
Scandinavia
Southern Britain
Orkney
French Basque
Germany
France
North Italy
Greece
Sardinia
Tuscany
0 100 200 300
b
Ireland
Scandinavia
Southern Britain
Orkney
French Basque
Germany
France
North Italy
Greece
Sardinia
Tuscany
Anatolia
0.0 0.4 0.8

## 2.7 Conclusion

### 2.7.1 Discussion of simulation model

The simulation model employed here is relatively complex compared to previous related human demographic / evolutionary models (Barbujani *et al.* 1995, Ray *et al*. 2003, Excoffier 2004, Currat and Excoffier 2005), as the inclusion of a selected allele and three distinct interacting cultural groups was necessary for the questions we were addressing. We adopted a statistical approach, ABC, which allowed us to estimate demographic and genetic parameters by integrating both genetic and archaeological lines of evidence – with no priority given to either. However, as with any simulation model of population history, many simplifying assumptions have to be made and the extent to which these assumptions may lead to erroneous conclusions remains unknown. For example, the model does not consider the *reverse-cause* hypothesis (Simoons 1970, Aoki 1986, Bayless *et al.* 1971, McCracken 1971a, 1971b) – which proposes that dairying first arose in populations that were already LP – because both ancient DNA evidence (Burger *et al.* 2007) and data from lipid residues on pots (Evershed *et al.* 2008) are inconsistent with this view. However, this does not mean that once LP-dairying gene-culture coevolution was established, conversion to the culture of dairying was more likely in high LP frequency populations. Such a process is captured in our model to an extent, in that 'cultural' conversion is determined by the frequency of the receiving cultural group (see section 2.4.4), and LP is unlikely to rise to high frequencies anywhere without the presence of dairying. Nonetheless, a more explicit treatment of this process may lead to different conclusions.

Some parameters, such as those relating to the effects of climate zone / elevation, and the logistic growth rate, are fixed based on realistic assumptions (Bellwood 2005, Colledge *et al.* 2004, Hassan 1981). For those parameters that are allowed to vary within a range we note that an important shortcoming is that in any single simulation their value is constant over the 360-generation duration of the run. This may be a particular issue for selection acting on an LP allele in $F_D$ (see below). Since we identify 'good' simulations using their fit to only two datasets – the farming arrival time and LP allele frequency at a range of geographic locations – it is unsurprising that our analysis is relatively uninformative for some parameters. However, inclusion of these parameters does serve to reflect uncertainty in their values.

Estimates for the arrival dates of farming at the 11 locations considered here were calculated as local weighted averages of calibrated carbon-14 dates (Pinhasi *et al.* 2005) from a Gaussian sampling region (see section 2.6.1). We set the standard deviation of this region at the mean nearest-neighbour distance to ensure that most of the carbon-14 data was used. However, the geographic density of carbon-14 dates is highly uneven across Europe and so the number of such dates that are informative for farming arrival time at any of the 11 locations will vary. Also, there appears to be a considerable amount of noise in the dates for the earliest farmers; for example, the earliest carbon-14 date for farming in Ireland pre-dates those for Great Britain, the Low Countries and Denmark. To test if these concerns had a major effect on our results, we reanalyzed our simulation data by using as observed farming arrival times the dates inferred by assuming a constant rate of spread of farming (estimated at 0.9 km/year (Pinhasi *et al.* 2005)) and calculating the great circle distance from Anatolia

to each sampling location. The results of this reanalysis were very similar to those presented in the previous section (see Itan *et al.* 2009 supplementary figures S3 to S6).

We are aware that the spread of the Neolithic over Europe was not as constant as our model assumes. After the arrival of the Neolithic in the Balkans, there is a pause of approximately 800 years before it starts to spread to central Europe, and there is another pause of 1,000 years before it spreads further into the northern German lowlands and other parts of northern Europe. Clearly, the carbon-14 dates we have used to estimate the farming arrival times will not fully reflect the complex history of Neolithization in all parts of the continent. Future improvements to this model will incorporate the effects of past vegetation, climate variation and other geographic features on migration parameters and carrying capacities. Further, integrating evidence from analyses of ancient and modern genetic diversity on the demography of domesticated cattle and other species – and, indirectly, their associated farmers (Troy *et al.* 2001, Bollongino *et al.* 2006, Edwards *et al.* 2007 and Chapter 6; section 6.5.3) – should provide a more complete picture of the Neolithic.

### 2.7.2 Discussion of results

The list of parameters for which the marginal posterior distributions are notably narrower than their corresponding prior ranges (selective advantage, intrademic gene flow, the mobility factors of $F_D$ and $F_{ND}$, and the geographic origin location of LP / dairying coevolution) – which we interpret as those parameters for which our analysis is informative – is an unsurprising one since we would expect

these parameters to have the greatest influence on the spread of an LP allele and farming in Europe.

The estimated selective advantage conferred by a LP allele (mode = 0.0953; 95% CI = 0.0518–0.159) is in good agreement with previous estimates for Europeans (0.014 to 0.19, Bersaglieri *et al.* 2004). However, it should be noted that: (1) this estimate is for selection only in dairying farmers, who make up just under half of the population that we simulate; and (2) we assume that selection is constant over time. It is possible that selection favouring LP has in fact been episodic and possibly spatially structured in different climate zones (Flatz and Rotthauwe 1973, Beja-Pereira *et al.* 2003, Simoons 1978, 1980, 2001, Bloom and Sherman 2005). Episodic selection would be difficult to model without additional information on when those episodes were likely to have occurred. But we reason that constant selection strength is a more parsimonious assumption in the absence of evidence to the contrary. If, as modelled here, dairying farmers made up less than half of the European post-Neolithic population then we would expect the real continent-wide selection values for LP to average less that half of what we estimate here. Such a range of selection values are, however, still consistent with previous estimates based on haplotype decay (Bersaglieri *et al.* 2004).

However, contrary to our expectations, we did not find that the presence of a positively selected LP allele in early dairying groups increases the unlinked genetic contribution to the modern European gene pool of people living in the region where LP-dairying coevolution started. The main reason for this is likely to be the relatively high inferred rates of intra- and interdemic gene flow between dairying and non-dairying farmers and between neighbouring demes, respectively, leading to a rapid

erosion of any demographic 'hitchhiking' of unlinked genomic regions. Additionally, we only track the genetic contribution of people living in and around the deme of LP / dairying coevolution from the inception of this process. Since it takes some time for the LP allele to rise to appreciable frequencies, any demographic 'hitchhiking' effect may become important only after the allele centroid has moved some distance away from its origin deme.

Another notable result was obtained when we compared the range of expected 13,910*T allele frequencies at different European locations – from simulations accepted at the 0.5% tolerance level – to those observed. While all observed values were within the 95% equal tail probability interval of the simulated values, many were somewhat offset from the modes. We interpret this as indicating that our model does not fully explain the distribution of the 13,910*T allele in Europe. One possible explanation for this is that migration activity – as modeled here by interdemic gene flow and unidirectional migration – has increased subsequent to the expansion of farming into the northwest reaches of Europe. In this scenario the farming expansion phase, occurring 9,000 to 5,500 ya, would be mainly responsible for generating the -13,910*T allele frequency cline in Europe but higher migration activity following this period would then have a homogenizing effect in LP allele frequencies. Intriguingly, a general pattern can be seen (figure 2.5b) whereby observed frequencies are lower than expected in northern Europe and higher than expected in southern Europe. Such a pattern is the opposite of what we would expect under the *calcium assimilation hypothesis*, which suggests higher selection for LP in northern latitudes through a greater requirement for dietary vitamin D and calcium because low-sunlight conditions reduce UV-mediated vitamin D production in the skin (Flatz and

Rotthauwe 1973). This frequently cited mechanism (Ingram *et al.* 2009, Simoons 1978, 1980, 2001, Weiss 2004, Hollox *et al.* 2001, Akey *et al.* 2004) was not included in our model and thus would seem to have negative explanatory power. Thus our simulations indicate that geographically and temporally homogeneous selection, in combination with well-attested underlying demographic processes, are sufficient to explain, indeed, to over-explain, the LP / latitude correlation in Europe. However, it should be noted that since we have not explicitly included a parameterized latitudinal effect on selection in our model, there may be scenarios where such an effect could also explain patterns of LP in Europe.

As inferred here, the spread of an LP allele in Europe was shaped not only by selection but also by underlying demographic processes; in this case the spread of farmers from the Balkans into the rest of Europe. We propose that this combination of factors could also explain the apparent homogeneity of LP-associated mutations in Europe. In Africa there are at least four known LP-associated alleles, including three that are likely to be of African origin (Ingram *et al.* 2007 Tishkoff *et al.* 2007) as well as -13,910*T, which is likely to be of European origin (Mulcare *et al.* 2004, Coelho *et al.* 2005). The greater apparent diversity of LP-associated mutations in Africa may reflect a greater genetic diversity in general, leading to the availability of more mutations upon which selection can act following the advent of dairying.

However, we suggest that this diversity is the result of an 'imposition' of dairying culture on a pre-existing farming people, rather than the spread of dairying being tied to the spread of dairyers. Such a model would require the availability of a number of, albeit low-frequency, LP-causing mutations; either through a high mutation rate or a large number of potential LP-causing sites. It is therefore possible

that, in the absence of the spread of dairying being linked to a major demographic expansion, high LP-allele diversity will also be found in the Indian subcontinent.

### 2.7.3 Dairy farming and the Linearbandkeramik (LBK) culture

Perhaps the most interesting result from our simulation model was the estimation of the location and timing of the start of LP / dairying coevolution. Our posterior modal estimates indicated it occurred in a region between the central Balkans and central Europe (figure 2.3), around 7,441 ya . These estimates appear consistent with a number of archaeologically well supported patterns concerning the emergence and spread of dairying. Recent carbon isotope ratios from lipids extracted from archaeological sherds show the presence of milk fats in present day western Turkey and connect these findings to an increased importance of cattle herding (Evershed *et al.* 2008, Röhrs and Herre 1961, Boessneck and Driesch 1979, Buttenhuis 1995, Benecke 1998). In general, the spread of the Neolithic lifestyle from the Aegean to central Europe goes hand in hand with the decline of the importance of sheep and goats and the rise in frequency of cattle bones in archaeological assemblages. While the Balkans at the beginning of the Neolithic still show a variety of subsistence strategies (Bartosiewicz 2005), the middle Neolithic in southeast Europe and the earliest Neolithic in central Europe after 7,500 ya show a clear preponderance of cattle. Benecke (1994a) gives the following averaged rates for the respective domestic species: cattle 55.2%, sheep and goats 32.6%, pigs 12%. The proportion of cattle in central Europe increases during the following centuries to an average of 73% and then stays (with a few exceptions) stable for most prehistoric

periods of Middle and northern Europe. Thereby, cattle herding is in most cases connected with kill-off profiles indicative for dairying (Bartosiewicz 2007, Benecke 1994a, Arbogast 1994, Balasse and Tresset 2002, Tresset 1996, 1997, Benecke 1994b).

Dairying, and the consumption of milk in particular, have been proposed to date back as far as the pre-Pottery Neolithic B of the Near East, and may even have driven the process of cattle and goat domestication (Cribb 1987, Vigne and Helmer 2007). While it was evidently a common cultural practice during all phases and regions of the European Neolithic, a fully developed dairying-based farming economy first emerges only during the late Neolithic in southeast Europe and the Middle Neolithic cultures following the Linearbandkeramik (LBK) in central Europe.

In the Mediterranean, milking of cattle occurs episodically (Vigne 2006) and sheep and goats remain the dominant domesticates, as they were earlier in Anatolia and the Aegean. It is very likely that the goat and sheep, and to a lesser extent cattle, based economies of the Mediterranean used processed milk in the form of yoghurt, cheese and other milk-derived products instead of fresh milk. The nutritional and agricultural differences between southern Europe, the Mediterranean and central- and northern Europe, as well as historic reports, point to this. For instance, the Romans used goat and sheep milk for the production of cheese, and cattle as a draught animal. In contrast, the Germanic peoples and other inhabitants of central and northern Europe practiced cattle dairying and drank fresh milk in significant amounts. The Greek historian Strabo (~64 BC to 24 AD) reports in his *Geography* (Strabo 1969): *"Their* ['the men of Britain'] *habits are in part like those of the Celti, but in part more simple and barbaric - so much so that, on account of their inexperience,*

*some of them, although well supplied with milk, make no cheese; and they have no experience in gardening or other agricultural pursuits."*

Overall, by considering the results from our simulations and archaeological, archaeozoological, and archaeometric findings, it seems very plausible to connect the geographic origin of the spread of LP to the increasing emergence of a cattle-based dairying economy during the 6th millennium BC. The geographic region of origin of the LBK – in modern day northwest Hungary and southwest Slovakia (Pavúk 2005, Bánffy 2004) – certainly correlates well with our results (see Itan *et al.* 2009 supplementary figure S7). The date of origin of LP / dairying coevolution estimated here (mode = 7,441 ya; 95% CI = 6,256 to 8,683 ya; see figure 2.4a and table 2.3) also fits well with dates for the early LBK in central Europe (~7,500 ya) and its proposed main predecessor, the Starčevo culture of the northern Balkan Peninsula and south of Lake Balaton (8,100 to 7,500 ya; (Baldia 2003)).

However, as explained above, our date estimate is conditioned by farming arrival dates in the estimated LP / dairying coevolution origin region. As a result, our date and location estimates are not independently derived. Nonetheless, a role for LP / dairying coevolution in the later rapid spread of LBK culture – from its origins in the Carpathian Basin – into central and northwest Europe would be consistent with the significantly higher mobility factor we infer for $F_{\mathrm{D}}$ compared to $F_{\mathrm{ND}}$. This is also consistent with the rapid dissemination of the LBK culture over a territory of 2,000 km width and approximately one million square kilometres within less than 500 years (Lüning 2005).

### 2.7.4 Concluding summary

In conclusion we infer that the coevolution of European LP and dairying originated in a region between central Europe and the northern Balkans around 6,256 to 8,683 ya. We propose the following scenario: after the arrival of the Neolithic in southeastern Europe and the increasing importance of cattle herding and dairying, natural selection started to act on a few LP individuals of the early Neolithic cultures of the northern Balkans. After the initial slow increase of LP frequency in those populations and the onset of the central European LBK culture around 7,500 ya, LP frequencies rose more rapidly in a gene-culture coevolutionary process and on the wave front of a demographic expansion (see Itan *et al.* 2009 supplementary videos S1 to S3), leading to the establishment of highly developed cattle-, and also partly goat-, based dairying economies during the Middle Neolithic of central Europe around 6,500 ya. A latitudinal effect on selection for LP, through an increased requirement for dietary vitamin D (Flatz and Rotthauwe 1973), is unnecessary to explain the high frequencies found in northern Europe.

# Chapter 3

# Late Pleistocene demography and the appearance of modern human behaviour[1]

## 3.1 Summary

The origins of modern human behavior are marked by increased symbolic and technological complexity in the archaeological record. In western Eurasia this transition, the Upper Paleolithic, occurred about 45,000 years ago, but many of its features appear transiently in southern Africa about 45,000 years earlier. I show that demography is a major determinant in the maintenance of cultural complexity and that variation in regional subpopulation density and / or migratory activity results in spatial structuring of cultural skill accumulation. Genetic estimates of regional population size over time show that densities in early Upper Paleolithic Europe were similar to those in sub-Saharan Africa when modern behavior first appeared. Demographic factors can thus explain geographic variation in the timing of the first appearance of modern behavior without invoking increased cognitive capacity.

---

[1] This is an expanded version of the article published as Powell A., Shennan S., Thomas M.G. (2009) 'Late Pleistocene demography and the appearance of modern human behavior' *Science* 324(5932): 1298-1301.

## 3.2 Abstract

I first outline the archaeological evidence for the appearance of modern human behaviour. This has long been associated with the Upper Palaeolithic in Eurasia and the Late Stone Age in Africa, but the far older evidence recently uncovered at Middle Stone Age sites across sub-Saharan Africa has made the geographic and temporal origins of modern behaviour less clear. I provide an overview of the explanations for the emergence of modern behaviour, before discussing how demographic arguments can account for the spatial and temporal heterogeneity in the appearance of modern behavioural markers.

I provide the details of a theoretical model of cultural evolution due to Henrich (2004a), which derives a relationship between a population's size and the level of culturally inherited skill that it can accumulate. Then I describe the development of our demographically explicit simulation model (section 3.4), which adapts and extends Henrich's model, making it applicable to investigating the cultural dynamics of human populations during the Late Pleistocene. This model demonstrates that spatial structuring of cultural skill accumulation is possible in environments with heterogeneous subpopulation density or migratory activity.

I then discuss the recent evidence of regional population expansions during the Late Pleistocene, obtained through Bayesian skyline analysis of a global dataset of mtDNA sequences. I use the resulting effective population densities to demonstrate that demography is a plausible explanatory mechanism for the heterogeneous appearance of the markers of modern human behaviour.

## 3.3 Introduction

### 3.3.1 Evidence for the appearance of modern human behaviour

The Upper Paleolithic (UP) transition, which occurred in Europe and western Asia about 45 thousand years ago (kya) (Mellars 2005, Bar-Yosef 2002), and later in southern and eastern Asia (Petraglia 2007, James and Petraglia 2005), Australia (Brumm and Moore 2005, O'Connell and Allen 2007), and Africa (Ambrose 1998b), is seen by many as marking the origins of modern human behavior. UP material culture, usually referred to as the Late Stone Age (LSA) in Africa, is characterized by a substantial increase in technological and cultural complexity, including the first consistent presence of symbolic behavior, such as abstract and realistic art and body decoration, e.g. threaded shell beads, teeth, ivory, ostrich egg shells, ochre, and tattoo kits; systematically produced microlithic stone tools, especially blades and burins; functional and ritual bone, antler, and ivory artefacts; grinding and pounding stone tools; improved hunting and trapping technology, e.g. spear throwers, bows, boomerangs, and nets; an increase in the long-distance transfer of raw materials; and musical instruments, in the form of bone pipes (Mellars 2005, Bar-Yosef 2002, Brumm and Moore 2005, Ambrose 1998b, d'Errico *et al*. 2003, McBreaty and Brooks 2000).

In Europe and western Asia, the UP transition happened relatively rapidly, with most of the characteristic features listed above appearing (the 'full package'), and is thought to coincide with the appearance of anatomically modern humans (AMH) in a region previously occupied by Neandertals (Zilhão 2007, Bocquet-Appel and Demars 2000, Mellars 2004, d'Errico 2003). In southern Siberia and northeast Asia,

microlithic technology appears between 43 and 27 kya (Brantingham *et al.* 2001, Hoffecker 2005), but a fuller UP package is not evident until ~22 kya (Brantingham *et al.* 2004). The evidence from south and southeast Asia and Australia also points to a more gradual accumulation of modern behavioral traits, such as ornamentation, use of ochre, and possibly rock art (Petraglia 2007, James and Petraglia 2005, Brumm and Moore 2005, O'Connell and Allen 2007). These are thought to first appear soon after the initial expansions of AMH into the regions but only become widespread later on, ~30 kya (James and Petraglia 2005) and ~20 kya, if not later (Haberle and David 2004, Brumm and Moore 2005), in south Asia and Australia, respectively.

In Africa the idea of a single transition has been contested (McBreaty and Brooks 2000, Henshilwood *et al.* 2002) because there is strong evidence for the sporadic appearance of many markers of modern behavior at multiple sites as early as 70 to 90 kya (Bar-Yosef 2002, McBreaty and Brooks 2000, Henshilwood *et al.* 2004), and possibly as far back as 160 kya (Marean *et al.* 2007). The African Middle Stone Age (MSA) sites of Katanda, Democratic Republic of Congo ~90 kya (McBreaty and Brooks 2000); Klasies River mouth (Howieson's Poort and Still Bay industries), South Africa ~65 to 70 kya (McBreaty and Brooks 2000, Jacobs *et al.* 2008); and, in particular, Blombos Cave, South Africa ~75 kya (Zilhão 2007, Henshilwood *et al.* 2002) present a striking array of modern traits, including the earliest evidence of abstract art (d'Errico *et al*. 2003, Henshilwood *et al.* 2002), as well as geometric blades (Lombard 2008), barbed bone harpoon points (McBreaty and Brooks 2000), bone awls, marine shell personal ornaments (Zilhão 2007), and engraved ostrich shell containers (Texier *et al.* 2010). However, these markers are intermittent and disappear between ~75 and 60 kya before making a more stable and widespread

reappearance in the LSA starting ~40 kya (Ambrose 1998b, Zilhão 2007, Henshilwood *et al.* 2004).

### 3.3.2 Explanations for the emergence of modern behaviour

Notwithstanding the oversimplifications made in the above outline, any adequate account of the emergence of modern behavior would need to explain not only the transition itself but also its heterogeneous spatial and temporal structuring (Bar-Yosef 2002) and earlier transient appearance in sub-Saharan Africa (McBreaty and Brooks 2000, Zilhão 2007, Henshilwood *et al.* 2002). Some authors have argued that behavioural modernity is a unique feature of AMH, usually invoking biological advances on our archaic ancestors; particularly those involving brain function as some have suggested (Klein 1999, 2000, Chomsky 2005). Genetic evidence for increased selection acting on human genes in the last ~40 k years (Hawks *et al.* 2007), faster rates of evolution for nervous system-related / neural genes (Dorus *et al.* 2004, Pollard *et al.* 2006), key changes in genes associated with language ability (Enard *et al.* 2002, but also see Krause *et al.* 2007) and signatures of recent selection acting on genes associated with brain development (Evans *et al.* 2005, Mekel-Bobrov *et al.* 2005, Gilbert *et al.* 2005, Evans *et al.* 2006, but also see Currat *et al.* 2006) provide intriguing candidates for biological change.

The precise nature of the AMH-specific brain changes remains open to speculation. Some have argued for increasing integration of a previously modular cognitive tool kit (Mithen 1996, see also Sperber 1996), while others have invoked specific biological innovation in linguistic ability and / or symbolic capacity (Klein

1999, 2000, Chomsky 2005, Bickerton 1990, Lieberman 2007). However, it is now widely accepted that AMH evolved in Africa ~160 to 200 kya (McBreaty and Brooks 2000, White *et al.* 2003, McDougall *et al.* 2005, Stringer 2007) and expanded into most habitable parts of the Old World between 90 and 40 kya (Lahr and Foley 1998, Ambrose 1998a, Ingman *et al.* 2000, Ray *et al.* 2005). If the primary cause of behavioral modernity was heritable biological change just prior to the UP / LSA, then any such mutation(s) would have had to rise to substantial frequencies after human populations had dispersed out of Africa; implying either their rapid spread around the world in the past 45,000 years or, potentially, geographic structuring of cognitive capacity. Furthermore, it is difficult to account for the early appearance of many modern behavioural traits in southern Africa some 70 to 90 kya (Zilhão 2007, McBrearty and Brooks 2000, Henshilwood *et al.* 2002), and their subsequent disappearance until the UP / LSA, with a biologically determined cognitive advance.

Many authors have argued that AMH (Mellars 2005, McBreaty and Brooks 2000, Zilhão 2007, Stringer 2007, Lahr and Foley 1998), and possibly even Neandertals (d'Errico *et al*. 2003, Conard 2005, Zilhão 2007), possessed the requisite capacities long before the UP / LSA. This raises the further question of why there was a delay of some 100,000 years between anatomical modernity and perceived behavioral modernity (Mellars 2005, Stringer 2007). Furthermore, while the full package of UP artefacts does not make an appearance in Australia until around 7 kya (Brumm and Moore 2005, O'Connell and Allen 2007) it would appear undeniable that the first Australians possessed advanced seafaring technologies when they colonised the continent some 33 to 53 k years earlier (O'Connell and Allen 2004, Hudjashov 2007).

A number of mechanisms triggering the expression of modern behavior have been proposed, many of which invoke demographic change as a causal factor. These include expansion into new environments necessitating the invention of new technologies (Stiner and Kuhn 2006), increased subpopulation density escalating intergroup resource competition (Mellars 2005, Stiner and Kuhn 2006) or social organization (Mellars 2005), increased intergroup interaction requiring various inter- and intra-group cultural signaling mechanisms (O'Connell and Allen 2007, Stiner and Kuhn 2006, Vanhaeren 2005), and increased stimulus for exoteric language, possibly leading – via the development of more complex recursive syntax – to the capacity for 'cognition by description' (Chomksy 2005, Bolender 2007, Russell 1988, also see Everett 2005).

### 3.3.3 Demography and cultural evolution

Two recent cultural evolutionary models (Shennan 2001, Henrich 2004a) explicitly demonstrate the positive effect of increasing population size on the accumulation of beneficial culturally inherited skills. These models support the proposal that demography was an integral component in the appearance of modern behavior (see also Stringer 2007).

The first, due to Shennan (2001), is a cultural drift model based on the work of Peck (Peck 1996, Peck *et al.* 1997), which showed that that when the results of cultural-innovation processes are passed on by a combination of vertical and oblique transmission, larger populations have a major advantage over smaller ones. When populations are small, innovations that are less reproductively beneficial and less

attractive to imitate are more likely to be maintained within them purely through sampling effects. Members of larger populations are on average both biologically fitter and more attractive as models for imitation as these deleterious sampling effects will decline as population sizes increase. Shennan's work (2001) was among the earliest to provide an explicit demographic model for the emergence of modern behaviour during the Late Pleistocene.

The second model (Henrich 2004a) draws on the growing body of work on the importance of biases in social learning (Boyd and Richerson 1985, see Chapter 1; section 1.2), and forms the basis for the work presented in the remainder of this chapter.

### 3.3.4 Henrich's model

Henrich's model (2004a) demonstrates that under certain critical conditions, directly biased transmission can lead to cumulative adaptation of a culturally inherited skill, even when the transmission process is inaccurate. Each individual in a population of size $N$ has a $z$-value $z_i$ that measures their level of ability at some cultural skill or in some cultural domain; such as the manufacture of tools, artefacts or weapons, the utilization of these technologies, or some kind of ecological knowledge. This value $z_i$ can be thought of equally as a quantitative measure of ability in some specific skill, or as a measure of the number of discrete skills or preferences possessed in a problem domain. Whilst under the first definition of this measure any increase in $z$-value would refer directly to an improvement in ability, under the second definition this would correspond to an increased breadth of cultural skills or

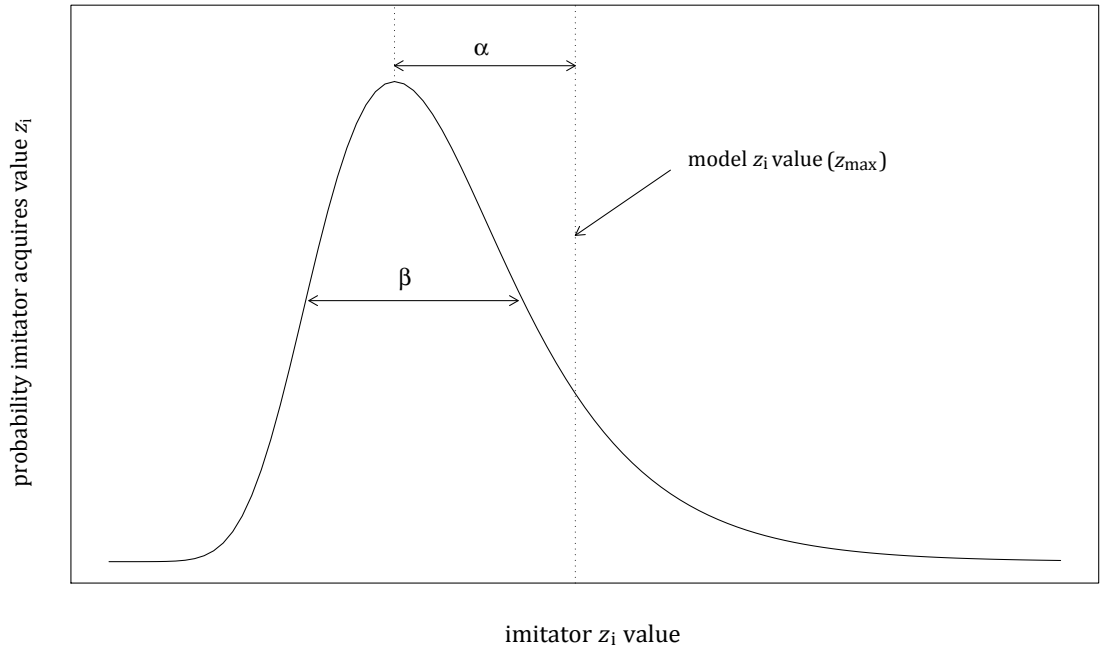
'toolkit' – either through invention *de novo* or the reapplication of a skill or preference from another domain. Members of this population attempt to learn this skill or toolkit by successfully identifying and learning from the maximally skilled individual (the 'cultural model'), i.e. direct bias (see Chapter 1; section 1.2).

In order to replicate an 'incomplete and inaccurate' process of inference, these learners never exactly replicate the *z*-value of their cultural model. Instead, any individual attempting to copy the model, with *z*-value $z_{\mathbf{max}}$ gains a value drawn at random from a Gumbel distribution with mode ($z_{\mathbf{max}} - \alpha$) and dispersion parameter $\beta$. This means that the transmission process is, firstly, systematically biased, as on average a copier will end up with a *z*-value less than that of their model by an amount $\alpha$ (the difficulty of inference, or 'complexity' of the skill or toolkit); and secondly, the process is 'noisy' so that there is a small probability, dependent on $\beta$ (which measures the degree to which copiers will make different inferences), that the copier will gain a *z*-value greater than that of their model.

Henrich utilized the Price equation to show that as population size *N* increases, the more likely it is that the positive combined effect of these occasional inaccurate inferences and the selective choice of cultural model to copy will outweigh the degrading effect of low-fidelity transmission. The Price equation (Price 1970, 1972, see also Frank 1995) provided a new way of accounting for change in any quantitative character in any evolutionary system. While Price's equation posits nothing new in terms of the evolutionary process, its novel partitioning of total evolutionary change into that due to its covariance with fitness, i.e. selection, and that due to stochastic transmission processes, i.e. drift, allowed existing theory to be seen in a new light. It

played an important part in demonstrating the theoretical equivalence between inclusive fitness theory (for example Hamilton 1970, 1975) and group-, or multi-level selection models (Chapter 1; section 1.2, see Okasha 2006 for a book length treatment). Brantingham (2010) provides a recent example of an application of the Price equation in evolutionary archaeology, which is able to delineate between potential selective and stochastic processes operating simultaneously on a population of cultural artefacts (see also Chapter 4; section 4.3).

**Figure 3.1:** A recreation of Henrich's transmission model (Henrich 2004a; figure 1). The probability of a learner gaining skill value $z_i$ from a cultural model with skill value $z_{max}$ is Gumbel distributed, with mode ($z_{\mathbf{max}} - \alpha$) and dispersion parameter $\beta$.



Henrich (2004a; appendix A, page 210) derived the following relationship giving the change in mean *z*-value per time step in terms of the transmission parameters α and β, and the population size *N*:

$$\Delta\bar{z} = -\alpha + \beta(\varepsilon + \ln(N))$$

where ε is the Euler-gamma constant ≈ 0.577. When this change in mean *z*-value is positive, i.e. $\Delta\bar{z} > 0$, the mean level of cultural skill in the population is increasing over time. Henrich terms this 'cumulative adaptive evolution', and rearranging the previous equation gives the necessary conditions:

$$N^* > e^{\frac{\alpha}{\beta} - \varepsilon}$$

where $N^*$ is the critical population size of social learners needed for an inferential / learning process specified by parameters α and β. Figure 3.2 shows this relationship, demonstrating that a larger population is able to maintain cultural skills that are more complex to learn, or larger cultural or technological toolkits.

Henrich applied this model to explain the apparent erosion of cultural complexity in Tasmania (Diamond 1978) throughout the Holocene following its separation from the Australian landmass 12–10 kya (Henrich 2004a, but see also Read 2006 and Henrich 2006). Recently Kline and Boyd (2010), in line with this

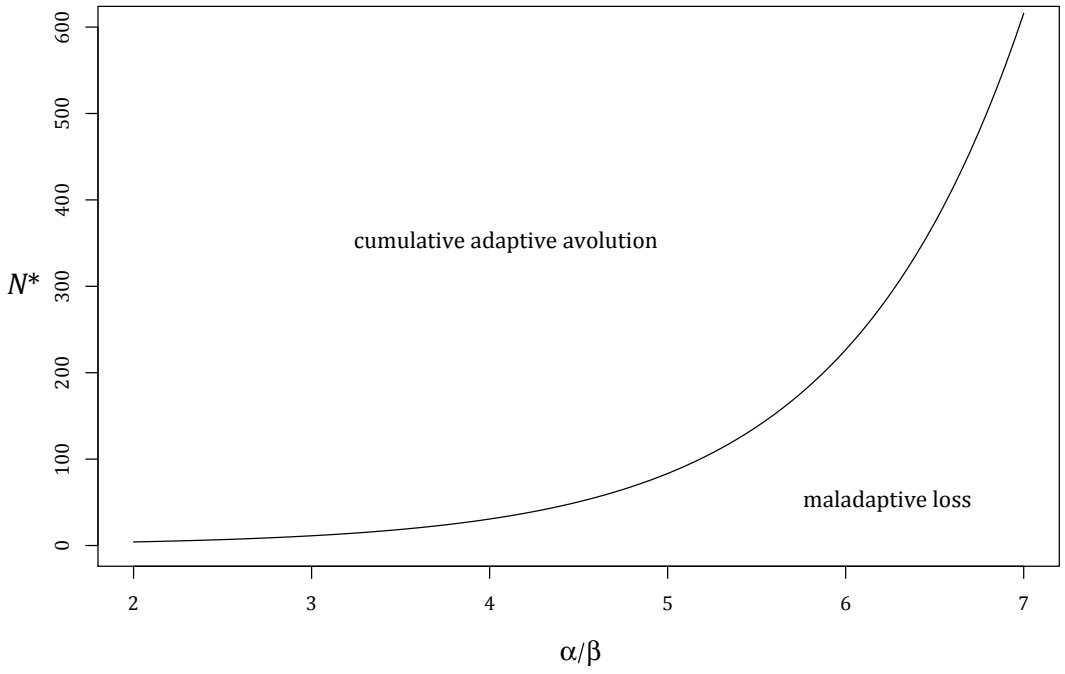
general model, have demonstrated that the complexity of technological toolkits in the island populations of Oceania at the time of European contact are correlated with population size and the level of inter-island contact.

## 3.4 The demographically explicit simulation model

Here, I adapt and extend Henrich's transmission model (2004) through simulation, into a more realistic structured metapopulation which reflects plausible Late Pleistocene conditions, to investigate the effects of demographic factors on the accumulation, or loss, of cultural complexity.

To apply Henrich's model three key unrealistic assumptions need first to be addressed: (a) All individuals belong to a single, large population; Late Pleistocene metapopulations, similarly to modern hunter-gatherers, would have comprised of small, highly dispersed subpopulations connected by migratory activity (Stiner *et al.* 2000, Stiner and Kuhn 2006); (b) Naïve individuals can accurately identify the most skilled member of the preceding generation as an oblique model; in large populations this would become impossible and, we suggest, even in small populations, identification of the most skilled adult will be inaccurate; (c) Naïve individuals always copy oblique models; cultural skills and behaviours are often first, and sometimes only, learned from parents (Shennan and Steele 1999, see also McElreath and Strimling 2008).

**Figure 3.2:** The relationship between the transmission process parameters α and β and the critical population size *N** necessary for cumulative adaptive evolution of a cultural skill. Recreated, in truncated from, from Henrich (2004; figure 2).




### 3.4.1 Simulation world

Our simulation model consists of *G* subpopulations placed at random in a 2-dimensional 'world' at a density *D*. Each subpopulation contains *N* adult individuals and is connected by local subpopulation density-dependent migratory activity. To reflect demographic conditions in the Late Pleistocene I set the subpopulation size *N* at 25, which, if we deem the skill to be inherited by same sex-offspring and that half the population are sub-adults, corresponds to actual group sizes of 100 – a

conservative value given various estimates of early human and contemporary hunter-gatherer population sizes (see for example Aiello and Dunbar 1993, Dunbar 1993, Birdsell *et al.* 1973).

### 3.4.2 Migration process

These subpopulations are connected by Gaussian random-walk migratory activity, with standard deviation $M_{sd}$ (the 'migratory range'), such that the mean global migration rate approximates the subpopulation density $D$. This is achieved by allowing all adult individuals to virtually undergo this 'walk' and if any 'hit' another subgroup they migrate to that group; otherwise they remain in their original subpopulation. $M_{sd}$ was defined as a proportion of the mean nearest neighbour distance $\bar{r}_E$ between subpopulations within the simulation world, calculated as $\bar{r}_E = 1/2\sqrt{D}$ (Clark and Evans 1954). Because ~99.7 percent of the probability density of a Gaussian distribution lies within ±3 standard deviations, when $M_{sd} \leq 1/3$ most subpopulations would be effectively isolated from one another. Preliminary test simulations – where I calculated the mean global migration rate – confirmed this and demonstrated that for $M_{sd} > \sim 0.33$, the mean global migration rate approximates the global subpopulation density $D$. Unless otherwise stated $M_{sd}$ is set well above this critical value at 1.0, i.e. equal to $\bar{r}_E$.

### 3.4.3 Cultural transmission processes

In order to augment the transmission process, I introduce a stochastic transmission model that incorporates both vertical and skill-level dependent oblique learning processes. In each generation of the model a naïve offspring generation also of size *N* is created in each subpopulation. Each offspring individual undergoes a process of 'vertical transmission', receiving a *z*-value according to the transmission process described by Henrich (2004) but from their parent rather than the maximally skilled individual. Then all offspring individuals are given the chance to undergo non-parental 'oblique transmission', by selecting an oblique model from amongst the subset of adults in their subpopulation with *z*-values greater than those they received from their parent, with probability proportional to the magnitude of the *z*-value difference. If no such 'qualified' oblique models exist for an offspring individual then this step is skipped. The offspring individual receives a *z*-value from the chosen oblique model, again through the transmission process as detailed above, but only retains this if it exceeds its current *z*-value.

### 3.4.4 Simulation process

In each generation of the simulation model the full algorithm is as follows:

(1) Create a generation of offspring of size *N* within each subpopulation.

(2) Each offspring individual chooses a parent by randomly sampling the adult generation with replacement.

(3) Naïve offspring undergo a process of 'vertical transmission', receiving a *z*-value from their parent according to the transmission process described by Henrich (2004).

(4) All offspring individuals are given the chance to undergo 'oblique transmission' as described above.

(5) The now 'encultured' offspring generation replaces the adult generation within each subpopulation.

(6) These new adults are then able to undergo local subpopulation density-dependent migration.

The critical population size $N^*$ in Henrich's result is dependent only on the ratio of $\alpha/\beta$ for a specific cultural skill. For ease of notation I collapse this ratio into a single parameter $\alpha$, which I term 'skill complexity'. Varying this composite parameter $\alpha$, which incorporates both the level of inferential noise and the systematic bias in the learning process, then allows simulation of the learning of cultural skills or toolkits of varying complexity.

All simulations are initialized by giving every adult in all subpopulations a *z*-value of 10.0 and then run forward for 100 generations. The mean level of cultural skill accumulation $\bar{z}$ is measured by averaging *z*-values across all individuals in all subpopulations. If the mean *z*-value in the final generation is greater than 10.0, i.e. $\Delta\bar{z} > 0$, then the result is deemed 'cumulatively adaptive'. To account for stochastic variation in simulation outcomes, I average results across 100 iterations.

The simulation model was written in the Python programming language (van Rossum *et al.* URL: http://www.python.org/), and post-simulation data was processed and visualized using the R language (R Development Core Team 2009, URL: http://www.R-project.org/).

## 3.5 Simulation model results

### 3.5.1 The number and density of subpopulations

I first explored the effects of varying the number of subpopulations $G$ in our simulated world on the mean level of cultural skill accumulated $\bar{z}$. For values of $G >$ ~50, the $z$-value did not increase much further across the entire range of subpopulation densities $D$ and skill complexities $\alpha$ (see figure 3.3). Figure 3.4 illustrates that the degree of skill accumulation increased with increasing subpopulation density and decreasing skill complexity. These results indicate that the accumulation, or maintenance, of culturally inherited skills is not dependent on the absolute metapopulation size, but rather on the degree of interaction of the constituent subpopulations given that $G >$ ~50. However, when $G <$ ~50, skill accumulation will, to an extent, be dependent on $G$, and thus the size of the metapopulation. This result may have some bearing on debate concerning the erosion of cultural complexity in Holocene Tasmania (Henrich 2004a). As a conservative measure, $G$ is fixed at 100 in all subsequent simulations.

**Figure 3.3:** The effect of increasing the number of subpopulations $G$ in the simulation world on mean $z$-value. Results are shown for a five different skill complexities $\alpha$ and four subpopulations densities $D$. Blue: $D = 0.001$; green: $D = 0.01$; red: $D = 0.1$; and black: $D = 0.15$.

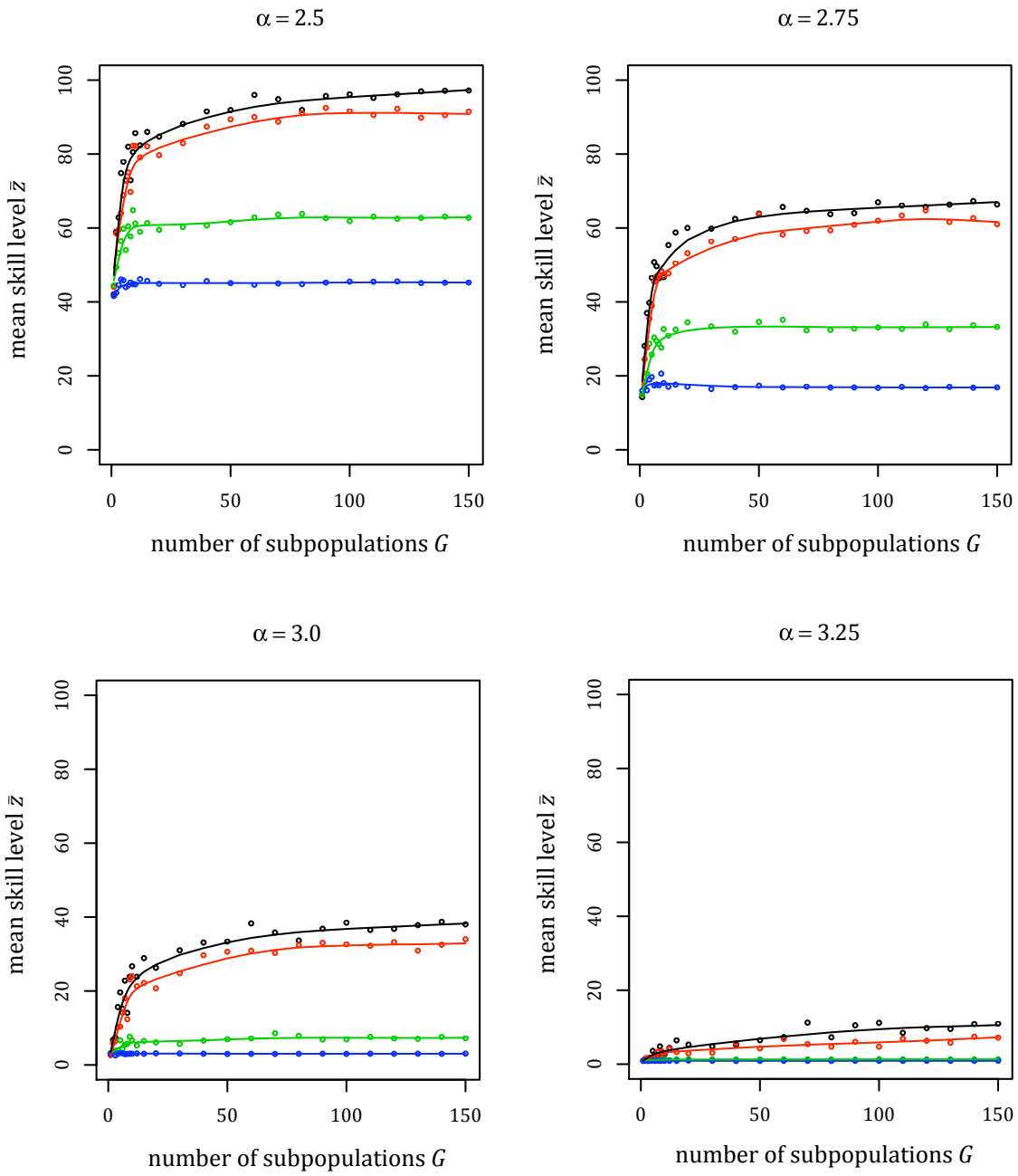
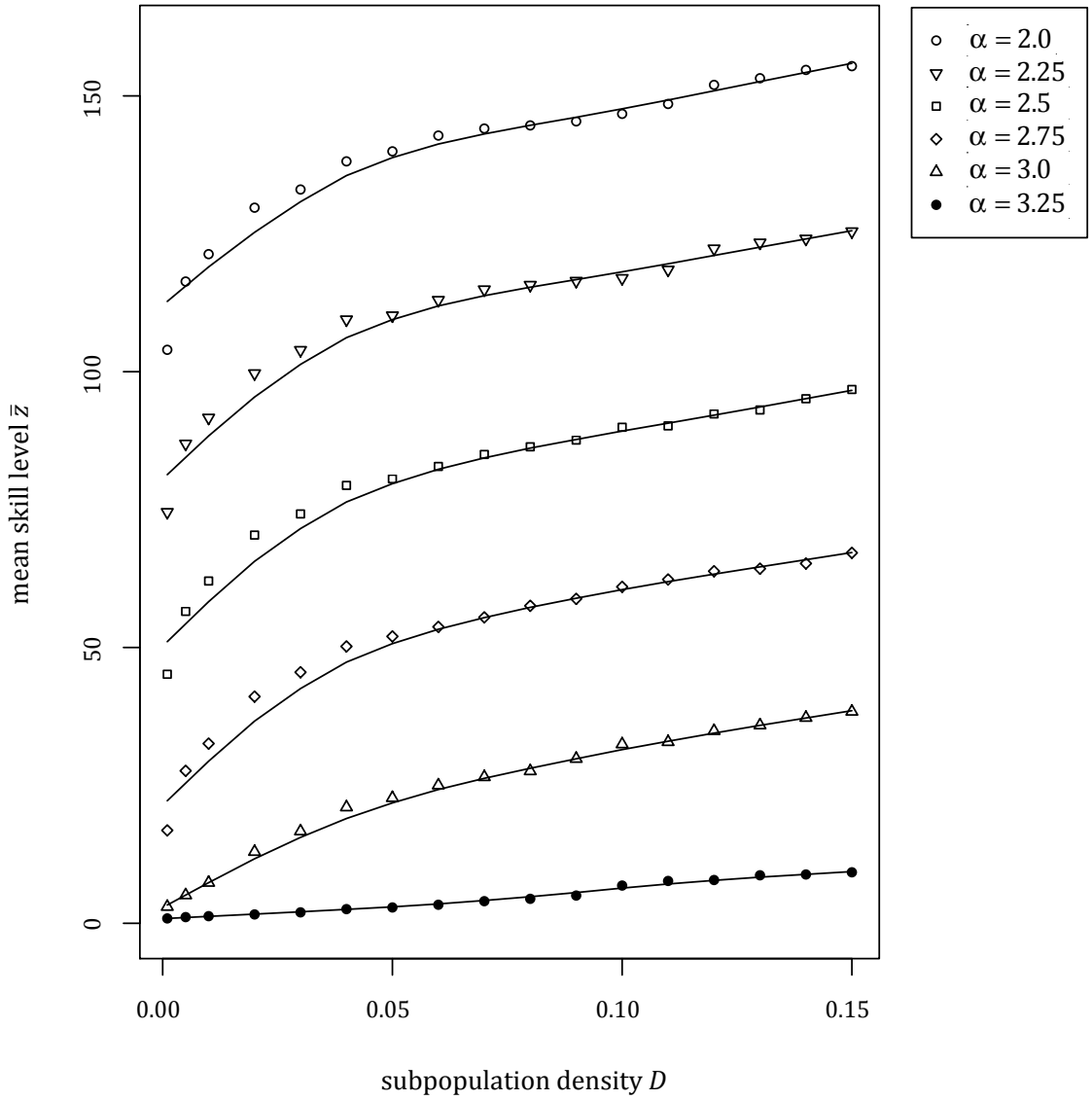


**Figure 3.4:** Mean *z*-values in the final (100th) generation, averaged over 100 iterations, for a range of values of skill complexity α and subpopulation density *D*.



### 3.5.2 Heterogeneity in subpopulation density *D*

A key feature of the UP is the geographic heterogeneity in apparent onset times, despite different regions being mutually accessible with modest migration activity. To investigate whether skill accumulation can be spatially structured as a result of different subpopulation densities, I partitioned the simulated world into two regions differing in density by an order of magnitude, $D_{\text{high}}$ and $D_{\text{low}}$. I retained $M_{\text{sd}}$ at 1.0, but as a proportion of the mean nearest-neighbour distance $\bar{r}_{\text{E}}$ in the lower density region, i.e. $\bar{r}_{\text{E}} = 1/2\sqrt{D}_{\text{low}}$. This ensured that sufficient subpopulations were connected by migratory activity – including across this virtual partition – for the migration rate to approximate the density in each region. I set $D_{\text{high}} = 0.02$ and $D_{\text{low}} = 0.002$ and simulated over a range of $\alpha$ values (2.0 to 4.0). For all $\alpha$ values, the skill accumulation was consistently higher in the $D_{\text{high}}$ region even though the two regions were contiguous. As an example, with $\alpha = 3.0$, this difference in mean regional *z*-values, averaged over 100 iterations, was maintained over the entire duration of the simulation (figure 3.5). Figure 3.6 provides an illustration from a single iteration of the spatial structuring of skill accumulation (see also movie S1 from Powell *et al.* 2009).

**Figure 3.5:** Regional mean *z*-values (averaged over 100 iterations) over 100 generations in a heterogeneous subpopulation density world with $\alpha = 3.0$. The 95% confidence intervals ($CI_f$) for each region are given as dashed lines.

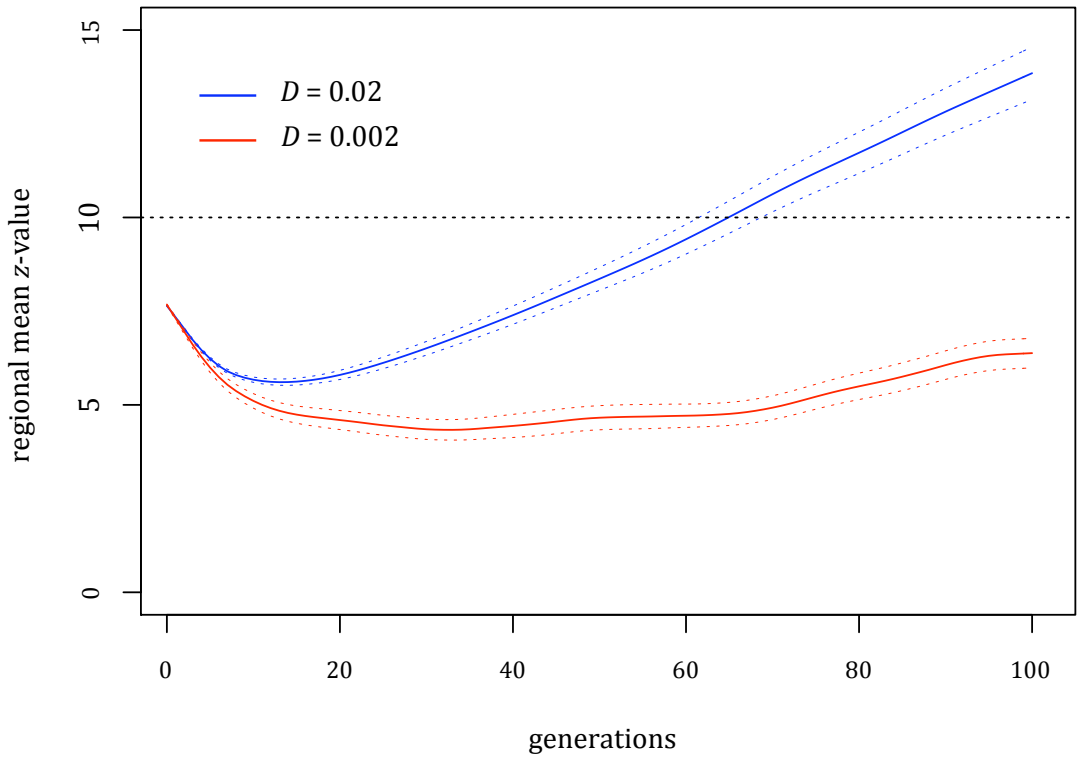
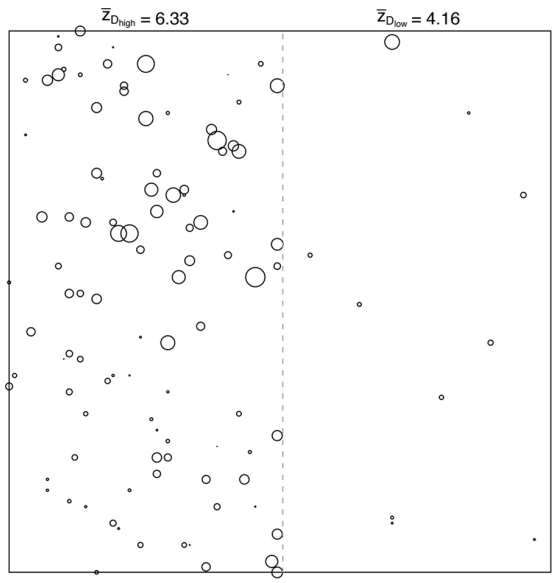
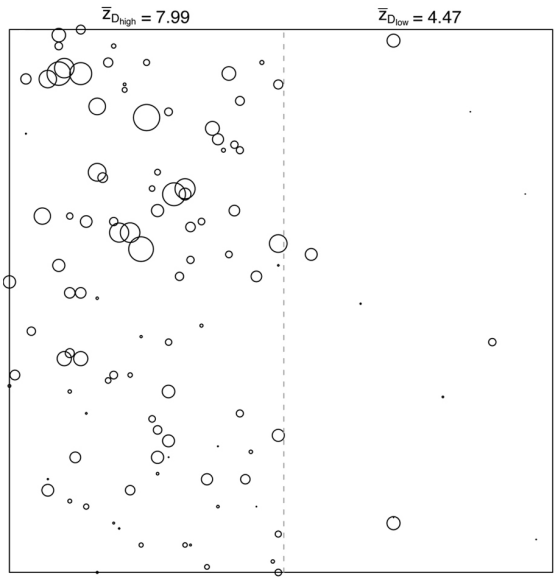


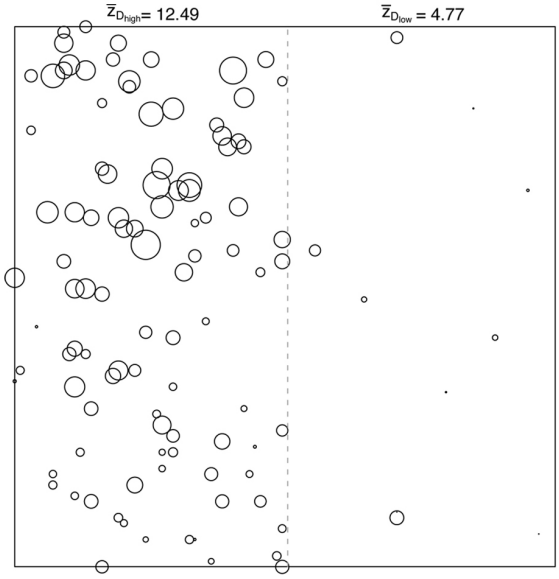

**Figure 3.6 (next page):** An illustration, from a single iteration with $\alpha = 3.0$ shown at 25-generation intervals, of the spatial structuring of skill accumulation in a heterogeneous subpopulation density world. The left side of each subplot is populated at density $D_{high} = 0.02$ and the right side at density $D_{low} = 0.002$. Each subpopulation is marked by a circle, centered on the spatial location of the group and with diameter proportional to its mean *z*-value. Regional mean *z*-values are also given at the top of each subplot.



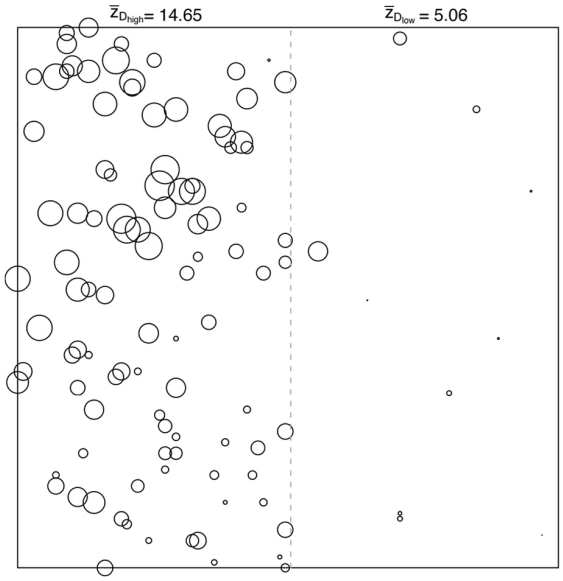

Gen 25



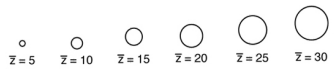

Gen 50




Gen 75




Gen 100



### 3.5.3 Heterogeneity in migratory range $M_{sd}$

In addition to regional density differences, we would also expect heterogeneity in migratory ranges during the late Pleistocene due to, for example, differing terrains, vegetation, or subsistence strategies (see James and Petraglia 2005). To investigate whether this could result in spatial structuring of skill accumulation, I populated the simulated world at a constant subpopulation density $D = 0.01$, and partitioned it into two regions with differing $M_{sd}$ values; $M_{sd,high} = 1.0$ and $M_{sd,low} = 0.1$, again allowing migratory activity across the partition. Similar to the heterogeneous density world, the skill accumulation was consistently higher in the well-connected $M_{sd,high}$ region across all $\alpha$ values simulated (2.0 to 4.0). An example, with $D = 0.01$ and $\alpha = 2.9$, is given in figure 3.7, with mean regional *z*-values averaged over 100 iterations. Figure 3.8 provides a spatial illustration from a single iteration (see also movie S2 from Powell *et al.* 2009).

**Figure 3.7:** Regional mean *z*-values (averaged over 100 iterations) over 100 generations in a heterogeneous migratory range world with $\alpha = 2.9$. The 95% confidence intervals for each region are given as dotted lines.

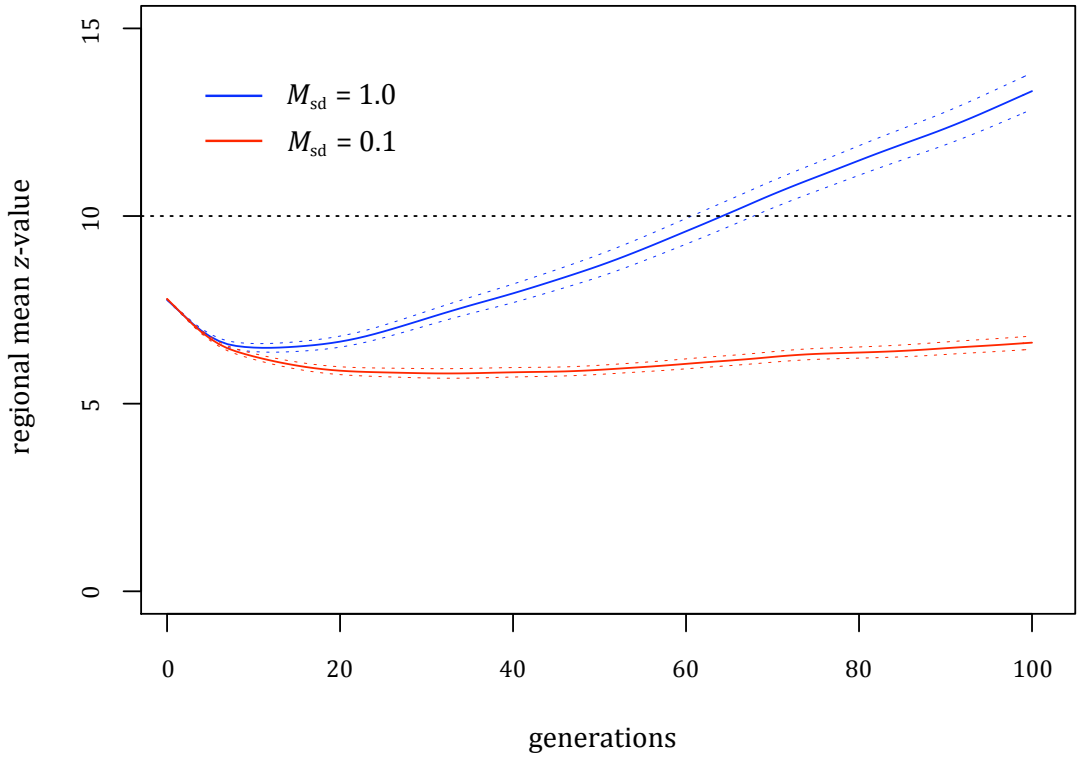
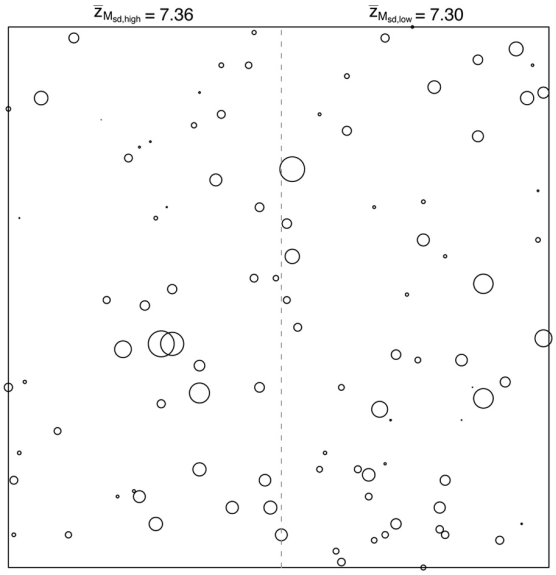
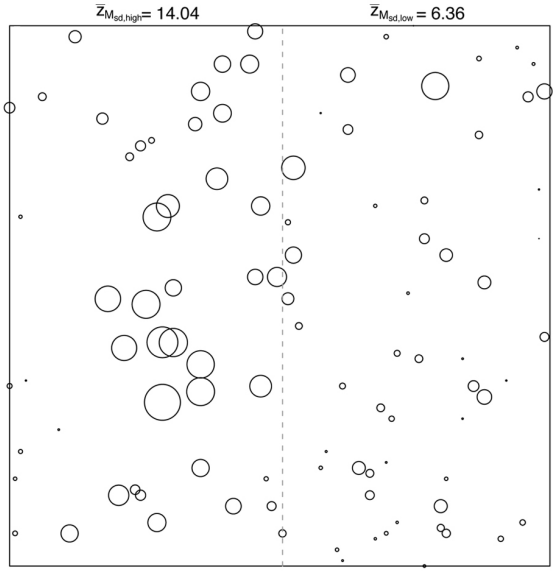


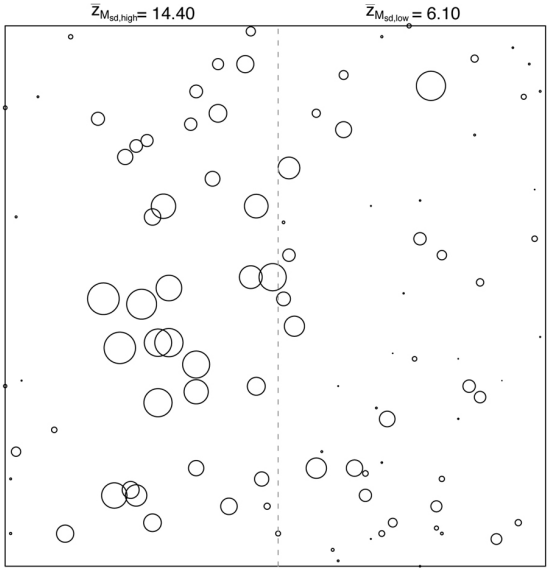

**Figure 3.8 (next page):** An illustration, from a single iteration with $\alpha = 2.9$ shown at 25-generation intervals, of the spatial structuring of skill accumulation in a heterogeneous migratory range world. Individuals in the left side of each subplot have migratory range $M_{sd,high} = 1.0$ and those on the right side have $M_{sd,low} = 0.1$. Each subpopulation is marked by a circle, centered on the spatial location of the group and with diameter proportional to its mean *z*-value. Regional mean *z*-values are also given at the top of each subplot.



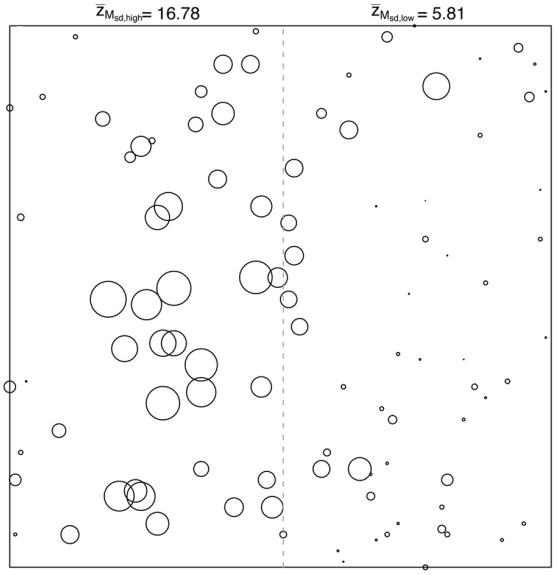

Gen 25



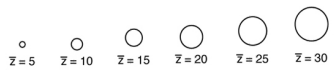

Gen 50




Gen 75




Gen 100

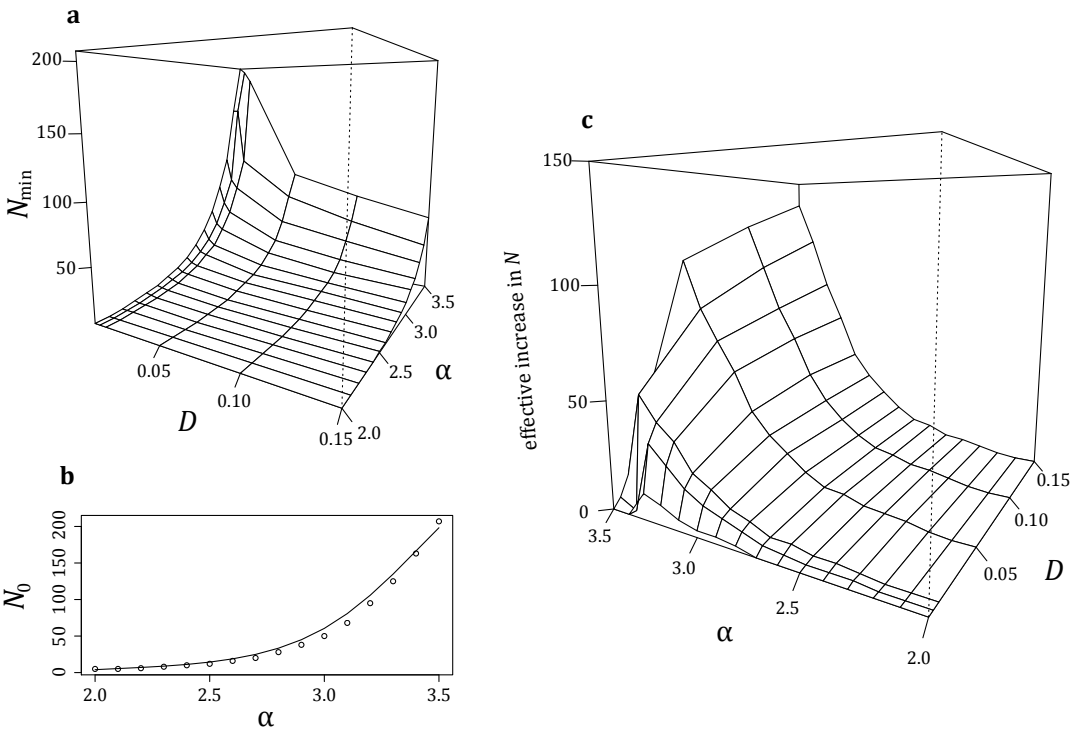


### 3.5.4 The effect of migration on cultural skill accumulation

From the above results, it is clear that migratory activity among a set of subpopulations can have the same effect on skill accumulation as increasing the size of a single population (Henrich 2004a). This is because it increases the within-group variance in skill levels $z_i$, which feeds the selective directly biased transmission process and offsets the eroding effect of low-fidelity transmission. I therefore sought to quantify the effect of increasing migration activity in terms of the effective number of adult individuals available as transmission models within each subpopulation.

To achieve this, I inverted the previous simulation process; for given values of $\alpha$ and $D$, I simulated widely over $N$ to find the minimum number of adults $N_{min}$ needed in each subpopulation for adaptive cumulative evolution to occur (figure 3.9a). I repeated this process for the same range of $\alpha$, but with no migratory process operating, to obtain the minimum number of adults required for skill accumulation in a single isolated subpopulation $N_0$ for each value of $\alpha$ (figure 3.9b). This is directly equivalent to Henrich's analytical result (2004; figure 2, and recreated in truncated form above as figure 3.2) but uses the extended transmission process presented in this model. I then calculated the effective increase in $N$ due to migratory activity by finding the difference between the $N_{min}$ expected for given values of $\alpha$ and $D$, and the $N_0$ expected for the same value of $\alpha$. As can be seen from figure 3.9c, not only does increasing migratory activity have the same effect as increasing the size of an isolated population, but also this effect is greater for higher skill complexities $\alpha$.

**Figure 3.9:** The effective increase in adult subpopulation size due to migratory activity. **a** The minimum number of adults required for adaptive cumulative evolution to occur $N_{\min}$ for a range of values for subpopulation density $D$ and skill complexity α. **b** The minimum number of adults needed in a single isolated population $N_0$ for the same range of α values. **c** The effective increase in adult subpopulation size $N$ due to migratory activity for this range of $D$ and α, calculated by subtracting surface **a** from curve **b** extended along the $D$ axis. The axes of **c** have been rotated for display purposes.



### 3.5.5 Simulation model conclusions

These simulation results demonstrate that the influence of demography on cultural transmission processes provides a mechanism to explain three key features of the emergence of modern behavior in the archaeological record: the early appearance, and subsequent disappearance, of many modern traits in Southern Africa 90 to 70 kya; geographical heterogeneity in the timing of the UP outside Africa; and the delay between the emergence of AMH as a species and the material expression of modern behavioral traits. If, as proposed here, demographic factors are fundamental in shaping the evolution of human behavior, then well-supported estimates of late Pleistocene regional population densities will be crucial to understanding the UP / LSA.

## 3.6 Late Pleistocene demography

### 3.6.1 Genetic evidence for regional population expansions

Early genetic / coalescent modeling analyses utilizing mitochondrial DNA (mtDNA) 'mismatch distributions' (Harpending *et al.* 1993, Sherry *et al.* 1994, Rogers 1995, see also Chapter 6) have indicated marked increases in human populations during the late Pleistocene. Consistent with the Recent African Origin model (for an early review see Stringer and Andrews 1988), these analyses – which relied on comparatively simple demographic models and potentially inaccurate mutation rate estimates – indicated a population expansion in sub-Saharan Africa ~80–100 kya, before later growth ~40–50 kya in Asia and Europe.

A recently developed inference method, Bayesian skyline analysis (Drummond *et al.* 2005, see also Pybus *et al.* 2000, Strimmer and Pybus 2001) allows the estimation of the effective size of a population back through time to the most recent common ancestor (MRCA) of a sample of contemporary and / or ancient DNA (aDNA) sequences. This Markov chain Monte Carlo (MCMC) coalescent method (see Chapter 6; section 6.4.4) is able to simultaneously estimate evolutionary parameters, such as the mutation rate, and the demographic history of the population, through a 'Bayesian skyline plot' (BSP). This approach has been implemented in the software package BEAST (Drummond *et al.* 2005, Drummond and Rambaut 2007), and has been widely used in a variety of contexts, including inference on the past demography of species using aDNA sequences (for example Shapiro *et al.* 2004, Barnes *et al.* 2007).

A recent study by Atkinson *et al.* (2008) used BEAST with a global data set of 357 coding-region mitochondrial DNA (mtDNA) sequences to infer ancestral effective human population sizes throughout the late Pleistocene in various regions of the world (see Atkinson *et al*. 2008; figure 2a). Their results indicated slow population growth in sub-Saharan Africa from the time of the sample MRCA ~143 – 193 kya, followed by more rapid expansions outside Africa starting ~50–70 kya, in line with previous studies (Atkinson *et al*. 2008; figures 1 and 2c). The estimated modal effective population sizes from this analysis allowed some comparisons of relative densities in different regions of the world over time.

### 3.6.2 Population density and the appearance of modern behaviour

By setting the UP transition in Europe at 45 kya (Mellars 2005, Bar-Yosef 2002), we can infer the critical effective population size, and therefore density, necessary for the accumulation of markers of modern behavior. Although this transition is closely associated with the initial colonization by AMH, the rapid rise in skill levels under favorable demographic conditions that are observed in our simulations indicates that cultural intensification would be largely insensitive to the time since first occupation. I assume that the habitable area of Europe would not have included most of Scandinavia, resulting in an estimated area of 8.883 million $km^2$. The median effective population size estimate in Europe at ~45 kya is 2,905, with 95% highest posterior density (HPD, see Chapter 2; section 2.5.4) interval of 280.4 to 15,933.9, giving an effective population density of $\sim 3.2714 \times 10^{-4} km^{-2}$. This value is a reasonable proxy measure of subpopulation density, assuming that subpopulation size is constant, that allows for comparison between regions.

The time at which this density would have been reached in sub-Saharan Africa (estimated area ~24.270 million $km^2$) is ~101 kya. Although this is a relatively crude date estimate, and ignores the importance of the likely large heterogeneity in population densities at the local level, it does correspond well with the first appearance of modern behavioral traits in the region (McBrearty and Brooks 2000 Henshilwood and Marean 2003). Furthermore, applying this estimation method to the 'Middle East and North Africa' region (estimated area ~13.588 million $km^2$) gives a date of ~40 kya at which the critical density is reached, relatively consistent with the first evidence of modern behavior in the Levant and northeast Africa (Bar-Yosef 2002, Ambrose 1998b).

In southern Asia, the predicted time for the UP transition, ~52 kya, considerably predates the first archaeological evidence for modern behavior at ~30 kya (Petraglia 2007). Similarly, the date estimate for northern and central Asia, ~40 kya, predates that of the first full UP site found at ~22 kya (Brantingham *et al.* 2004). One possible explanation lies in the choice of the regions used in the analysis presented by Atkinson *et al.* (2008). An important assumption of the ancestral population size estimation method used is that samples are taken from unstructured, i.e. randomly mating, populations. Although multiple loci clustering analysis performed by Rosenberg *et al.* (2002; figure 1) (377 autosomal microsatellites for 1056 individuals from 52 populations) and more recently by Li *et al.* (2008; figure 1) (650,000 SNP loci for 938 individuals from 51 populations) broadly supports this assumption for most of the other regions, it clearly does not for either the 'southern Asian' or the 'north and central Asian' geographic regions Atkinson *et al.* (2008) created. Using this method of inference on such ancestrally structured data sets is likely to have resulted in an overestimation of the effective population size and / or the time at which population expansion took place. In addition, coalescence date estimates for major mtDNA haplogroups in southern Asia have been interpreted as reflecting an initial phase of population growth somewhat later (Sun *et al.* 2006) and even as recently as ~35 kya (Kivisild, personal communication).

A second possible explanation for this anomalous result is that, although population density may have been sufficiently high for behaviorally modern traits to otherwise accumulate, the migratory range may have been insufficient to allow wide-scale interaction between subpopulations. This may have been the case in southern Asia during the later Pleistocene (James and Petraglia 2005): *"The fragmentation of*

*viable ecological niches and the resulting isolation of populations from one another must have affected the transmission of cultural and technological practices."*

Although the inferred population densities (Atkinson *et al.* 2008) could account for the early appearance of behaviorally modern traits in sub-Saharan Africa and the Middle East – given our demographic model of cultural skill accumulation – they cannot explain the subsequent absence of these features between 70 and ~40 kya because no population size reduction during this period is inferred (Atkinson *et al.* 2008; figure 1). However, the method of coalescent inference used (Bayesian skyline analysis) may be unable to accurately reconstruct more complex demographic histories when using sequences sampled from a single locus (Drummond 2008), so repeated bottlenecks and / or expansions, which would have an important bearing on the accumulation of culturally inherited skills (Stringer 2007), may not be recaptured. Paleoclimatic data does indicate worsening conditions during oxygen isotope stage 4 (OIS4) ~75 to 60 kya, possibly leading to population decline, fragmentation, and range contractions during this period (Lahr and Foley 1998, Ambrose 1998a). Lahr and Foley (1998) suggest that continent-wide secondary population bottlenecks may have occurred ~70 kya, and there is some evidence that the sites of the South African Howieson's Poort industries became effectively depopulated by ~60 kya (Jacobs *et al.* 2008, Lahr and Foley 1998). In addition, we note that the population size growth for sub-Saharan Africa – as inferred from genetic data (Atkinson *et al.* 2008) – is gradual over the last 100 thousand years, whereas it is more dramatic in the rest of the Old World ~50 kya. Thus we should expect these African populations to be more sensitive to any climatic deterioration in terms of the accumulation of culturally inherited skills (Stiner and Kuhn 2006).

## 3.7 Conclusion

Our model provides a plausible explanation for the spatial and temporal structuring of the markers of modern behavior in the paleoanthropological record, even if all AMH had the requisite biologically determined cognitive capacities from the time of origin some 160 to 200 kya.

The simulation model presented here is conservative, in that we would expect a degree of positive feedback between population density and the accumulation of culturally inherited skills; the development of more advanced technologies, and possibly social organization, would likely lead to population growth. Furthermore, we would expect to see more artefactual evidence of behavioural modernity in higher population density contexts through increased deposition. Although the model we have presented does not accommodate these processes or explain the necessary cognitive developments that make possible the invention or improvement of complex behavioural traits, it does provide a demographic mechanism for limiting the degree to which early human populations would have accumulated these culturally inherited skills over time.

### 3.7.1 Neandertal culture

This chapter, and the paper on which it is based, avoids much of the complication involved in assessing the archaeological evidence for modern behaviour in Neandertals (for a review see d'Errico 2003), which in some cases is interpreted somewhat controversially (see Zilhão *et al.* 2006, Soressi and d'Errico 2007, Mellars *et al.* 2007). However, the recent publication of the draft Neandertal genome (Green

*et al.* 2010) indicates a more complicated relationship between this species and our own than some previous genetic analyses had allowed (Krings *et al.* 1997, Currat and Excoffier 2004, Orlando *et al.* 2006, but also see Nordborg 1998, Plagnol and Wall 2006).

Comparison of this reconstructed genome with those of present-day individuals indicates a greater similarity with individuals in Eurasia than in Africa, and suggests a degree of gene flow between the two species after expansion from Africa but prior to the fragmentation of the non-African population as it spread across Eurasia. While the apparent degree of interbreeding may be inflated by the demographic phenomenon of 'allele surfing' (Green *et al.* 2010, see also Edmonds *et al.* 2004 and Chapter 2; section 2.3), the fact that it appears to have occurred may have significant implications for assessing the level of inter-species cultural exchange that took place over the ~10 ky duration of cohabitation in Europe.

Some authors have argued that the Châtelperronian site in central France from ~35 kya represents the beginnings of an independent Neandertal shift towards behavioural modernity, possibly driven by contact with AMH (Zilhão and d'Errico 1999, d'Errico 2003, Zilhão *et al.* 2006), while others maintain that it is the result of, possibly repeated, interaction and 'acculturation' between the two species (see for example Mellars *et al.* 2007). Either way, the capability demonstrated in the production of lithic and decorated bone tools and the use of pigment, would suggest few, if any, cognitive limitations, and it seems likely that a demographic explanation – analogous to the argument presented in this chapter – would provide a greater understanding of the dynamics of Neandertal culture.

# Chapter 4

# The neutral model of cultural evolution: a test for neutrality

## 4.1 Summary

By borrowing a body of work from population genetics on the frequency dynamics of neutral alleles – the 'neutral model' (e.g. Kimura and Crow 1964, Kimura 1983) – evolutionary archaeologists, following Neiman (1995), have been able to test hypotheses concerning cultural variant frequency changes in various assemblages. This neutral or 'cultural drift' model, in which cultural variant frequencies are proposed to evolve analogously to those of neutral alleles, has in some cases provided a consistent explanation of observed diversity. However, in other contexts the observed diversity may allow this 'null' neutral model to be rejected, with various processes of cultural bias or selection then being invoked as causative.

In this chapter I provide an overview of the application of the neutral model in archaeology and its relation to power-law distributions of variant frequencies (e.g. Bentley *et al*. 2004). I then provide, through simulation, a novel means to robustly test for departure from the neutral model. I describe how this method is applied to a large cultural dataset (Schauer 2008), and is further extended to analyze the models of frequency-dependent copying biases introduced by Mesoudi and Lycett (2009). The results for the latter show that data generated under models with relatively high levels of copying bias cannot be rejected as being generated by a neutral process.

## 4.2 Abstract

I first provide a derivation of Neiman's neutral model (Neiman 1995) – based on the Wright-Fisher model of population genetics – which gives two expressions for the level of cultural diversity maintained in a population. The first, $t_F$, is an empirical estimator, based on observed cultural variant frequencies; while the second, $t_E$, is theoretically derived under the assumptions of neutrality and cultural drift / innovation equilibrium (section 4.4.1). Differences between these two statistics is interpreted as evidence for departure from neutrality, and I give archaeological examples of this. I then detail how the distribution of variant frequencies generated under the neutral model can be described by a power-law, and again provide examples of this approach applied to various cultural datasets (sections 4.4.3 and 4.4.5).

However, without a clearer understanding of how these statistics behave under neutrality for a range of different population sizes and innovation rates, it is difficult to reject the neutral model as generating an observed dataset. Section 4.5.2 describes how null distributions of these statistics can be empirically generated through simulation, and section 4.5.3 shows how they can be utilized to analyze a very large dataset of decorated Greek pottery.

The final section (4.6) applies this approach to the frequency-dependent copying bias models analyzed by Mesoudi and Lycett (2009). My results show that surprisingly high levels of copying bias acting within an otherwise neutral process can produce variant frequency distributions indistinguishable from those generated under complete neutrality. This demonstrates that population-level aggregation can

disguise the action of *non-random* processes to some extent, and that power-law distributions should not be interpreted too strongly as evidence for a neutral cultural process.

## 4.3 Introduction

Archaeologists have long been concerned with interpreting the individual elements of an assemblage as part of a related sequence, most commonly in terms of chronology. This method of arranging a series of artefacts, or 'seriation', relies on identifying stylistic similarities across successive stages – famous early examples include the work of Evans (1850), Pitt Rivers (1875) and Petrie (1899). Robinson and Brainerd (Robinson 1951, Brainerd 1951) used this approach with ceramic assemblages, but focused on the sample *frequencies* of various different styles through time. This method of 'frequency seriation' gives rise to the widely recognized 'battleship' curves, which track the changing sample frequencies of cultural styles, or 'variants', through time – characteristically reaching a maximal frequency 'peak' before declining.

The important work of Dunnell (1978, 1980) drew the focus of inquiry onto the population of artefacts themselves and, indeed, considered the study of this population as the most profitable in providing explanations for change in the archaeological record (see also Lipo *et al*. 1997). This perspective relies on the assumption that a sequence or 'lineage' of artefacts is causally linked through time by cultural transmission processes, i.e. through 'heritable continuity' (see O'Brien and Lyman 2000, Lipo *et al.* 1997). Dunnell's work (1978, 1980) also emphasized an important dichotomy between 'stylistic' and 'functional' attributes (see Shennan and Bentley 2007, and Shennan 2008 for further clarification on this terminology), and the marked difference cultural evolutionary processes would have on their frequencies in a population. 'Stylistic' variants are those whose frequencies would

fluctuate stochastically due to random 'noise' or 'error' in social transmission; whereas the frequencies of 'functional' variants would respond to directional cultural selection / social transmission pressures through preferential copying. While debate on the validity of such a strict dichotomy continues (for example see Brantingham 2007 and references therein) many analyses, including much of the rest of this chapter and Chapter 5, proceed on the assumption that in many contexts (e.g. Neiman 1995, Bentley *et al.* 2004) cultural variation can, in the absence of evidence to the contrary, be treated as stylistic i.e. neutral (but see Steele *et al.* 2010).

With this kind of 'population thinking' approach (Mayr 1982) to cultural assemblages, many evolutionary archaeologists (e.g. O'Brien and Lyman 2000 and 2002, Shennan 2002, and Shennan 2008 for a recent review) have recently begun to adopt a number of evolutionary models and techniques, including some directly analogous to those in genetics.

Some authors have attempted to identify cultural lineages from assemblage / frequency seriation data (O'Brien and Lyman 2000, 2002, 2003, Bettinger and Eerkens 1999, Lipo *et al.* 1997, Eerkens and Lipo 2005, Collard *et al.* 2006). Characteristics of these lineages are used to construct cultural phylogenies, which may then be interpreted as evidence for a particular population history or dynamic, in a similar way to those in evolutionary linguistics and anthropology (see for example Gray and Jordan 2000, Gray and Atkinson 2003, Gray *et al*. 2009, Holden and Mace 1997, Mace and Holden 2005, Tehrani and Collard 2009).

A second approach, pioneered by Neiman (1995), utilizes the Wright-Fisher neutral drift model from population genetics, where 'stylistic' cultural variant

frequencies are treated as analogous to neutral allele frequencies. Under the assumption that a drift / innovation equilibrium is reached (where the loss of cultural variants through drift is equal to the introduction of new variants through innovation), a measure of the expected level of cultural diversity can be obtained. This can then be used to test hypotheses on the cultural processes that gave rise to observed assemblage data. This approach was further extended and applied to a wide range of datasets by a number of other authors (Shennan and Wilkinson 2001, Brantingham 2003, Hahn and Bentley 2003, Bentley and Shennan 2003, Herzog *et al.* 2004, Kohler *et al.* 2004, Bentley *et al.* 2004, Schauer 2008, Steele *et al.* 2010). The neutral model approach may prove useful in explaining otherwise anomalous observations (e.g. Neiman 1995, Shennan and Bentley 2007), and forms the basis of the rest of this chapter – beginning with a full derivation of the model in section 4.4.

Another approach – related to the previous in that it is also firmly grounded in 'population-thinking' borrowed from genetics (Mayr 1982) – is that of Brantingham (2007, 2010). This approach makes no *a priori* assumptions on the neutrality or otherwise of cultural variants and uses the Price equation to derive a means of delineating the effects of drift and cultural selection operating simultaneously (Brantingham 2010). The Price equation, while most commonly applied to genetic evolution, provides a way of partitioning the total change in any evolutionary system into selective and stochastic components (Price 1970, see also Frank 1995) and has also been applied in other cultural contexts (see Henrich 2004, Powell *et al.* 2009 and Chapter 3).

A further body of work (e.g. Enquist and Ghiranda 2007, Enquist *et al.* 2008, Strimling *et al.* 2009) – while not explicitly focused on archaeological questions –

derives relationships between the size of a population and expected cultural diversity, in order to account for the vast 'accumulation of culture' that so uniquely characterizes our species (see for example Tomasello 1999). Much of this work (Enquist and Ghirlanda 2007, Enquist *et al.* 2008, Strimling *et al.* 2009, Ghirlanda *et al.* 2010) uses replicator dynamics to model social learning subject to biases, such as conformity. Under the assumption of cultural equilibrium these models are used to derive predictions about the 'amount of culture' that can be maintained in a population (Enquist and Ghirlanda 2007, Strimling *et al.* 2009, Ghirlanda *et al.* 2010).

## 4.4 The neutral model of cultural evolution

### 4.4.1 Neiman's neutral model in archaeology

Neiman (1995) presents a derivation of the neutral cultural model, which I first summarize in terms of the Wright-Fisher model of population genetics, before providing an overview of a number of its applications to archaeological datasets. The model utilized by Neiman is derived directly through analogy from a large body of work in population genetics on the evolution of neutral alleles (see for example Kimura and Crow 1964, Crow and Kimura 1970, Kimura 1983). The genetic 'neutral model' assumes that all possible alleles are selectively neutral (or very nearly neutral, see Ohta 1973), so that their frequencies in the population change solely through the stochastic processes of 'genetic drift' and mutation. Genetic drift is the process whereby allele frequencies change as a result of chance differences in reproduction, and can be simply considered as 'sampling error' in a finite population. One further

assumption is that any genetic mutations give rise to completely novel alleles in the population. An important corollary of this 'infinite alleles' assumption is that any exact identity between individuals *has* to be due to common descent. This is because mutations can only cause genetic differences between individuals to *increase* as back-mutations, which could disguise the true degree of relatedness through descent, are disallowed.

The Wright-Fisher model assumes a randomly mating *haploid* population of effective size $N_e$, which gives rise to – and is completely replaced by – a subsequent offspring generation also of effective size $N_e$. The probability of two randomly chosen offspring individuals sharing the same parent is $1/N_e$ and thus the probability that they do not is $1 - 1/N_e$. However, the probability that they share an *allele* through common descent is obviously not constrained solely to the previous generation, as they could share a common ancestor any number of generations ago. This probability $F_t$ – also known as homogeneity – is therefore defined recursively for generation $t$ as:

$$F_t = \frac{1}{N_e} + \left(1 - \frac{1}{N_e}\right) F_{t-1}$$

where $F_{t-1}$ is the similarly calculated homogeneity in the previous generation $t-1$ (Kimura and Crow 1964, but note the factor of 2 difference for their *diploid* case). This, of course, currently ignores the effects of mutation introducing novel alleles into the population. If we assume that mutation occurs at rate per generation $\mu$ then the probability that *neither* of our two randomly chosen individuals carry a mutation is

$(1 - \mu)^2$. Taking this into account the probability of identity by descent, or homogeneity, in generation $t$ is now:

$$F_t = \left[\frac{1}{N_e} + \left(1 - \frac{1}{N_e}\right)F_{t-1}\right](1-\mu)^2$$

Equilibrium homogeneity $\hat{F}_{eq}$ is reached when the loss of alleles through drift is exactly counter-balanced by the introduction of new ones through mutation. This means that once equilibrium is reached there should be no further change in homogeneity over subsequent time-periods. So by setting $F_t = F_{t-1}$ and rearranging the above it can be shown (Kimura and Crow 1964, but see also Shennan and Wilkinson 2001) that this equilibrium homogeneity is given by:

$$\hat{F}_t = \frac{(1-\mu)^2}{\left[N_e - (N_e - 1)(1-\mu)^2\right]}$$

Under the assumption that the rate of mutation per generation $\mu$ will be small, any $\mu^2$ terms will be negligible and can safely be ignored. So a good approximation of the equilibrium homogeneity is given by:

$$\hat{F}_{eq} \approx \frac{1}{2N_e\mu + 1}$$

This last equation makes it clear that the equilibrium homogeneity in a population is inversely proportional to the product of its effective size $N_e$ and the rate

of mutation $\mu$. While this result has been derived in terms of 'common descent' and 'homogeneity' in a population, it is more common – and perhaps more intuitive – to talk in terms of 'heterogeneity' or 'diversity', which is simply the reciprocal of homogeneity. This reciprocal, or 'diversity estimator' $\theta$, which gives the effective number of alleles a population can maintain at equilibrium, is thus:

$$\theta = 2N_e\mu$$

and therefore a larger effective population is able to support a higher level of genetic diversity for a given mutation rate. The diversity estimator $\theta$ is a parameter fundamental to population genetics (see Chapter 6; section 6.4.3), although in the more usual diploid case $\theta = 4N_e\mu$ as the population now contains *twice* the number of chromosomes as it does individuals.

Neiman (1995) draws his perceptive analogy with this population genetic model by considering the lines of descent as lineages of artefacts causally linked by a process of cultural transmission amongst their makers (O'Brien and Lyman 2000, Lipo *et al.* 1997). The homogeneity measure $F_t$ now gives the probability of two randomly chosen artefacts having the same cultural variant due to common descent through a cultural transmission chain. Following the 'infinite alleles' assumption that insists that all mutations produce novel alleles, Neiman assumes that all innovation events introduce completely novel cultural variants into the population. Neiman's diversity estimator $\theta$ is now considered as a measure of cultural diversity, giving the 'effective number of cultural variants' that can be maintained at equilibrium.

Neiman further describes how this diversity estimator $\theta$ can be estimated empirically directly from variant frequencies, an estimator he calls $t_F$. The empirical homogeneity statistic $\hat{F}$ can be calculated as:

$$\hat{F} = \sum_{i=1}^{k} p_i^2$$

where $p_i$ is the relative frequency of variant *i* of the *k* different variants in an assemblage. By assuming this assemblage data has arisen from a process that has reached equilibrium, a simple rearrangement gives an expression for this empirical diversity estimator:

$$t_F = \frac{1}{\sum_{i=1} p_i^2} - 1 \qquad (4.1)$$

The diversity estimator $\theta$ can also be estimated theoretically through the use of Ewens' sampling formula for neutral alleles (Ewens 1972). This formula gives the expected number of different alleles $E(k)$ in a sample as a function of the population size *n* and genetic diversity $\theta$, given the assumption of selective neutrality:

$$E(k) = \sum_{i=0}^{n-1} \frac{\theta}{\theta + i} \qquad (4.2)$$

Given an observed genetic sample which contains $k$ different alleles, a maximum likelihood estimate of $\theta$ can be obtained by iteratively incrementing or decrementing its value until $E(k)$ equals $k$. Neiman simply replaces 'alleles' with 'cultural variants', and terms the resulting theoretical estimator of cultural diversity $t_E$ (Neiman 1995).

Neiman goes on to note the fact that this theoretical estimator $t_E$ may be adversely affected by sample size; larger assemblages may result in a deflated estimate. Neiman explains this as a consequence of applying Ewen's sampling formula – which describes the genetic diversity in a population at equilibrium at a *instantaneous* point in time – to archaeological assemblage data that is likely to have accumulated over very long periods of time, i.e. over many 'cultural generations' (Neiman 1995; page 17). Neiman further suggests that $t_F$ estimates may be also be constrained in smaller assemblages, and that any significant departure of the *difference* between the two diversity estimators ($t_F - t_E$) from zero could be used to detect the existence of these potential confounding sample size effects in any analysis.

Another result that is absolutely critical to this approach was previously demonstrated by Neiman (1990; pages 182-193), and concerns the extent to which the *observed* assemblage variant frequencies are representative of the *true* frequencies in the underlying population of artefacts. Through simulation Neiman was able to show that the frequencies in assemblages produced by multiple transmission and discard events *would* closely match those in the whole population, even under a large range of discard rates.

Neiman applied this neutral model approach to explain the observed temporal variation in ceramic decoration in the Woodland Period assemblages in Illinois. First, he demonstrated that the difference ($t_F - t_E$) in this dataset was not statistically significantly different from zero, and therefore concluded that there was no untoward sample size effect. In line with demic genetic models (see Neiman 1995; page 21), Neiman hypothesized that elevated intergroup transmission between the five assemblage sites would increase effective population size and thus increase the level of within-assemblage diversity that could be maintained. Making the reasonable assumption that the level of intergroup transmission would be constrained by physical distance he proceeded to show that, as expected, the amount of within-assemblage stylistic diversity was inversely correlated with inter-assemblage distance. Neiman thus concluded that the neutral processes of drift and innovation were sufficient to explain the observed diversity of the Woodland assemblages.

### 4.4.2 Further applications of the neutral model in archaeology

Further work with this model by Shennan and Wilkinson (2001) analyzed a dataset of ceramic band-types from the early (~7,300 ya) Neolithic Linearbandkeramik (LBK) settlements in western Germany (see also Chapter 2; section 2.7.3). However, in this approach Shennan and Wilkinson were interested in using the difference between the two diversity estimators ($t_F - t_E$) to test whether the distribution of LBK band-type frequencies deviated significantly from that expected under the neutral model. Recall that the diversity estimator $t_F$ is empirically calculated directly from sample frequencies, while the other $t_E$ is theoretically derived

under the assumption of neutrality (and also, implicitly, equilibrium). First they demonstrated that sample size effects of the kind discussed by Neiman could be disregarded; both diversity estimators $t_F$ and $t_E$ were negatively correlated with assemblage size, so diversity was not being constrained by smaller sample sizes (see Shennan and Wilkinson 2001; pages 585-586). Therefore they argued that any residual difference could be interpreted as evidence for some non-neutral process of cultural transmission. Then, while the diversity estimates from the early phases were consistent with what they would expect under neutrality, they demonstrated that in the later phases, the empirical estimator $t_F$ becomes significantly larger than the theoretical one $t_E$. This suggested that a larger level of diversity was being maintained than could be expected under neutrality, and they interpreted this as evidence of a pro-novelty or 'negative frequency-dependency' (Boyd and Richerson 1985) copying bias in the later phases. Deviations between the diversity expected under neutrality and the true sample diversity could, of course, have a number of different causes; an important one is the implicit assumption of equilibrium in the calculation of $t_E$. A non-equilibrium system may transiently support more or less diversity than would be expected even though the underlying processes are still neutral. The requirement for equilibrium in a cultural system is addressed in Chapter 5, where a model is developed that relaxes this assumption.

Kohler and colleagues (Kohler *et al*. 2004), demonstrated that a ceramic assemblage from a Late Coalition period village on the Pajarito Plateau, New Mexico was less stylistically diverse than would be expected under neutrality. They suggested that this reduced diversity is evidence of some kind of frequency-dependent conformist- (Boyd and Richerson 1985), or even prestige- (Henrich and Gil-White

2001) bias in social transmission. This may be explained by the fact that this period (~1250–1325 AD) is characterized by large community size increases in the area; conformity may therefore have played an important part in maintaining social cohesion during the rapid societal changes that followed (Kohler *et al.* 2004; page 114).

### 4.4.3 The neutral model and power-laws

A related 'neutral' approach to analyzing cultural datasets is that of Bentley and colleagues (see for example Bentley *et al*. 2004, and section 4.4.4 for more detail). They use the fact that at equilibrium the variant frequency distribution generated from such a neutral random-copying process can be approximated by a power-law distribution. This family of 'right-skewed' or 'long-tailed' distributions is characterized by an abundance of variants at low frequency and very few variants at extremely high frequency (more detail is given below). They are of the form $P(v) \sim v^{\alpha}$, where $P(v)$ describes the proportion of variants in the population at frequency $v$, and result in a straight line on a log-log plot with slope $\alpha$ (see figure 4.1 for an example). This approximation arises from the fact, demonstrated by Kimura and Crow (1964), that at equilibrium in a finite population the variant frequency distribution is given by $P(v) = \theta(1-v)^{\theta-1}v^{-1}$. The power-law approximation $P(v) \sim v^{\alpha}$ holds when the innovation rate $\mu$, and thus $\theta$ (= $2N_e\mu$) is low, with the exponent / slope $\alpha$ varying as a function of $\theta$ in the underlying neutral process. By fitting the variant frequency distributions from a wide range of different cultural contexts to a power-law distribution on log-log axes, Bentley and colleagues are able to make inferences on

the parameter $\theta$, and with independent estimation of the effective population size $N_e$ this allows estimation of the innovation rate $\mu$ (see Herzog *et al.* 2004 for example). More detail on this work is provided in section 4.4.5. Schauer (2008) attempts to combine the previously outlined diversity estimator approach pioneered by Neiman (Neiman 1995, Shennan and Wilkinson 2001) with this frequency distribution / power-law approach with some success and, again, more detail of this work is provided below.

### 4.4.4 A short digression on power-laws

Power-law distributions of the form $P(v) \sim v^{\alpha}$ described above have long been recognised in abundance across a wide range of both natural and man-made phenomena, and have collected a rather daunting number of names and suggested mechanisms for their generation (see Newman 2005; section 4). The fundamental underlying property that distributions of this form share is that of 'scale-invariance'. This means that scaling the dependent variable (e.g. variant frequency $v$) by a constant factor $c$ simply scales the function (e.g. proportion of variants at a specific frequency $P(v)$) by the constant $c^{\alpha}$ and leaves its shape unchanged. It follows that *any* ratio of variant proportions is a constant over the whole range i.e. $P(cv)/P(v) \sim (cv)^{\alpha}/v^{\alpha} = c^{\alpha}$. This relationship holds for whatever value of $v$ we look at, and therefore does not depend on the scale.

This inverse relationship between rank / size and frequency / abundance has been recognised by scholars in many disparate fields. One commonly used name is

'Zipf's law', after the American linguist George Kingsley Zipf (Zipf 1929, 1949), though this refers specifically to distributions that have exponent equal to –1, i.e. of the form $P(v) \sim v^{-1}$. Another often used name is the 'Pareto distribution', named after the Italian economist Vilfredo Pareto (Pareto 1896), which is actually a cumulative distribution function, so in the context of this chapter it would refer to the proportion of variants with frequencies *larger* than a given frequency $v$.

Such power-law distributions have been found to describe a wide range of phenomena. Some examples are: the magnitude and frequency of earthquakes (Gutenberg and Richter 1944), asteroid size (Dohnanyi 1969) and thus moon crater diameter (Neukum and Ivanov 1994), the sizes of cities (Rosen and Resnick 1980, Gabaix 1999), the distribution of wealth in human societies (Pareto 1896), word usage frequencies in languages (Zipf 1929, Simon 1955), the number of biological species within a taxa (Willis and Yule 1922) and their rate of extinction (Newman and Eble 1999), the sizes of US firms (Axtell 2001) and *their* rate of extinction (Cook and Ormerod 2003, Ormerod 2005), and even the distribution of scientific citations (Redner 1998, Simkin and Roychowdhury 2006). Many further examples are given in Sornette (Sornette 2006) and Newman (Newman 2005) and, indeed, a comprehensive online bibliography of power- and Zipf law papers dating from as far back as 1881 is maintained by Wentian Li at http://www.nslij-genetics.org/wli/zipf/.

The appearance of these scale-invariant distributions has often been described as an 'emergent' large-scale outcome of an aggregation of myriad small-scale processes, and has led to the development of many models which generate such population level distributions from plausible small-scale interactions. One such class of explanatory models has a long history, and is based on the idea that the quantity or

entity of interest increases or grows in proportion to its current value or size; in other words 'the rich get richer'. Two early examples of this type of process were described by Yule (1925) and Simon (1955), hence the name: the Yule-Simon distribution, which is a type of stochastic urn process. Yule was attempting to provide a mechanism to explain the observed power-law distribution of the sizes of biological genera in terms of the number of species they contain. To summarize, Yule assumed – perhaps over-simplistically – that as existing species provide the 'raw material' for potential new species, then the number of new species that arise within a genus would be in proportion to the number already present. Simon essentially re-derived the Yule distribution (Simon 1955; page 427) and provided further examples, but using the theory of stochastic processes that had emerged during the 30 years since publication of the original result.

This model is now more commonly known as 'preferential attachment' (Barabási and Albert 1999), and has been utilized extensively in investigating the formation and structure of large-scale networks. Rather than being completely *random* networks, such as those first described by Erdős and Rényi (1960), a large number of real-world networks appear to be better described by the model of 'preferential attachment'. This model assumes that networks grow by the continuous addition of new vertices, which preferentially attach to existing vertices dependent on their current level of connectedness. From this process arises a 'scale-free' (or 'invariant') network where the connectivity of its vertices is power-law distributed, meaning that the network can critically depend on a few extremely well-connected 'hubs'. Preferential attachment can be contrasted with the formation of random networks, where edges between each pair of vertices are formed with a constant

small probability. This results in a Poisson distribution for the connectivity of each node (Barabási and Albert 1999; page 510), meaning that the likelihood that vertices with *k* connections exist decreases exponentially as *k* increases; so comparable well-connected 'hubs' are rare or often absent.

Some famous examples of real-world scale-free networks include: the structure of the World Wide Web (Barabási and Albert 1999), the web of sexual partners in a Swedish behavioural survey (Liljeros *et al.* 2001), human dispersal, using bank-notes as a proxy measure (Brockman *et al.* 2006), and the structure of metabolic networks (Jeong *et al.* 2000). Uncovering the common structure of such systems may provide for more effective policy in combating the spread of contagious diseases, or addressing vulnerabilities such as the targeted disruption (or even removal) of critical communication or travel network 'hubs'.

However, Frank (2009) describes an alternative basis for the seeming ubiquity of power-law distributions based solely on information theory. Frank demonstrates (2009; page 1571-1572) that any large-scale aggregation process that preserves information solely about the geometric mean is simply 'attracted' towards a power-law distribution, so that invoking a *specific* generative model may be redundant. Indeed, he suggests that these proposed models may provide minimal explanatory power, if any: *"The consistency of a generative mechanism with observable pattern tells us little about how likely that generative mechanism was in fact the cause of the observed pattern"* (Frank 2009; page 1581). This does allow, however, for the creation and *rejection* of generative models that are *inconsistent* with the observed data, and this idea underlies my work presented in this chapter (sections 4.5 and 4.6).

### 4.4.5 Power-laws and neutrality in cultural datasets

Especially relevant to this chapter are a number of papers by Bentley and colleagues (see Bentley *et al*. 2007 for a comprehensive review), who interpret power-law distributions in various cultural datasets as being explicable in terms of an underlying neutral or 'random-copying' process.

Hahn and Bentley (2003) analyzed the 1,000 most popular male and female baby-names in the US during the 20th century, a period encompassing ~300m births in total. In each decade the popularity of the top 1,000 names were power-law distributed, a relationship that held over three orders of magnitude of popularity (i.e. from 0.01% to almost 10%) and, much to the probable dismay of parents everywhere, can seemingly be very well-described by frequency-dependent random copying.

Bentley and Shennan (2003) reanalyzed the abovementioned LBK dataset (Shennan and Wilkinson 2001), this time utilizing the framework of stochastic networks to test different hypotheses of cultural transmission. They showed that modelling the cultural transmission of ceramic styles as a scale-free network, analogous to a random-copying neutral model (as briefly outlined above), provided a *reasonable* power-law fit to the data. However, this band-type data were *better* described by a log-normal distribution, meaning a smaller proportion of variants at very low frequencies than would be expected under neutrality. The authors demonstrated that removing the condition of continuous network growth – to more accurately reflect the known population history of the region – but keeping the cultural transmission process neutral, produced well-fitting distributions of this kind. This is consistent with the earlier finding of an apparent *divergence* from neutrality in

the form of a pro-novelty copying bias (Shennan and Wilkinson 2001), but suggests a different causative mechanism and, indeed, one that may still be considered neutral in terms of transmission.

In a study of late 20th century US patent data, Bentley *et al*. (2004) showed that the number of citations received by patents containing certain keywords in either the title or main-text were power-law distributed. The *overall* citation distribution had too few patents cited extremely rarely to be power-law distributed over the entire range, but *was* so distributed provided the number of citations was at least ~10. Apart from this deviation, which may be due to legal / regulatory necessity, these results again suggest that at the population level patent citations can be well explained by the random-copying model.

A further study considered the changes in popularity of dog breeds in the US over a 55 year period (Herzog *et al.* 2004), and found that the number of new dogs registered with the American Kennel Club in each of the 100+ recognized breeds – around 42m dogs in total – was power-law distributed. They concluded that, contrary to industry opinion, the popularity of the overwhelming majority of dog breeds could be very well described with this random copying model. The one exception was an anomalous, yet short-lived, spike in the popularity of Dalmatians in the late 80s and early 90s, most likely a *non-random* response to the 1985 rerelease of the wildly successful Disney film *101 Dalmatians*.

These examples demonstrate that the neutral model may in many cases serve as a reasonable and parsimonious null model, from which exceptions can be identified and more complicated generative mechanisms invoked. However, no

robust statistical tests for deviation from neutrality have been used in any study discussed up to this point; this is the focus of the following section.

## 4.5 Testing for departure from the neutral model

### 4.5.1 Motivation

None of the neutral model studies previously outlined offer robust statistical justification for assuming or rejecting the neutral model. There is no measure of quite how 'good' the fit to a power-law, as measured by $r^2$, would have to be to satisfy the null neutral model. Without an idea of the distribution of this statistic under the null model, it is impossible to test for significant deviation away from neutrality in a dataset. Similarly, while Shennan and Wilkinson (2001) take a significant difference between the two diversity estimators ($t_F - t_E$) as evidence for a departure from neutrality, without a null distribution of this difference over a range of different combinations of population size and innovation rate it is very difficult to reject neutrality.

Given this, I decided to produce empirical distributions of both diversity indices $t_F$ and $t_E$, and the degree to which the variant frequency spectrum fits to a power-law ($r^2$), through repeated simulation under the neutral null model for a wide range of values for the effective population size $N_e$ and innovation rate $\mu$. These neutral, or 'null', distributions could then be used as the basis for hypothesis testing with real data, with population sizes and innovation rates estimated directly from the data. If the $r^2$ value for a specific dataset, with estimated parameters $N_e$ and $\mu$, is below the

$m^{th}$ percentile of the appropriate simulated null distribution, i.e. with the same $N_e$ and $\mu$ parameter values, we can reject neutrality at the $m$% level – see figure 4.4 below as an example. Similarly, if the diversity estimator difference $|t_F - t_E|$ is greater than the $(100 - m)^{th}$ percentile of the null distribution we can, again, reject neutrality at the $m$% level.

### 4.5.2 Null distributions of test statistics under the neutral model

To this end I created a neutral / random-copying model based on that of Bentley *et al.* (2004), which produced simulated null distributions of the three statistics of interest: $t_F$, $t_E$, and $r^2$, as well as the absolute difference $|t_F - t_E|$, for a wide range of values for the effective population size $N_e$ and innovation rate $\mu$ (see also Schauer 2008; page 39).

Following Bentley and colleagues (Bentley *et al.* 2004), the model is initiated with a population of $N_e$ individuals, each of which is characterized by a unique cultural variant i.e. at maximum heterogeneity. The simulation then proceeds in a series of discrete time-steps, or generations. In each generation a new offspring population of size $N_e$ is introduced, and each offspring individual gains a cultural variant by copying a randomly chosen parental individual i.e. through a drift process. However, with a small probability $\mu$ (<<1), which we call the 'innovation rate', an offspring individual innovates and gains a completely novel variant. So on average $\mu N_e$ offspring individuals gain a novel variant through innovation, and the remaining

$(1 - \mu)N_e$ gain their variants unchanged from the previous generation. This offspring generation then completely replaces the previous parental generation.

In the final generation the total numbers of copies of each individual variant are accumulated over all generations of the simulation. These variant frequencies are logarithmically binned and normalized by scaling by the size of the bin (see Bentley and Shennan 2003). The *proportion* of variants that fall into each of these copy number bins is then calculated and this is plotted against the mid-point of each bin. This variant frequency distribution is fitted to a power-law model and the $r^2$ value of this fit is calculated. $t_F$ is calculated as described in equation 4.1 above, and $t_E$ is iteratively approximated using equation 4.2 also given above. $|t_F - t_E|$ is also calculated and recorded. This simulation process is iterated 1,000 times, after which we construct empirical distributions, and tabulate percentiles, of the three statistics: $t_F$ , $t_E$ and $r^2$ and the absolute difference $|t_F - t_E|$.

I repeated this process for a wide range of values for effective population size $N_e$ and innovation rate $\mu$, where $100 \leq N_e \leq 8{,}000$ and $0.01 \leq \mu \leq 0.1$. An example of the results from a *single* iteration of the simulation is given in figure 4.1, with population size $N_e = 500$ and innovation rate $\mu = 0.02$.

In order to reduce simulation time, which increases rapidly as $N_e$ and $\mu$ increase, I terminate each simulation soon after equilibrium is reached – i.e. where the opposite effects of drift and innovation on variant diversity have become equalized. This is achieved by running two separate iterations simultaneously; the first of which (A) starts with every individual having a unique cultural variant i.e. at maximum heterogeneity; and the other (B) is initialized with just one variant in the

population i.e. at maximum homogeneity. Both simulations are run forward in parallel and the model calculates $t_F$ in every generation in both A and B. When $t_F^{(A)} - t_F^{(B)} < 0$, i.e. when the heterogeneity of B becomes greater than that of A, we know that equilibrium is likely to have been reached (see figure 4.2) and thus halt this pair of iterations after allowing a further 20 generations in order to be conservative.

**Figure 4.1:** An example of the variant frequencies generated in one iteration of the neutral model, and the resulting power-law fit.

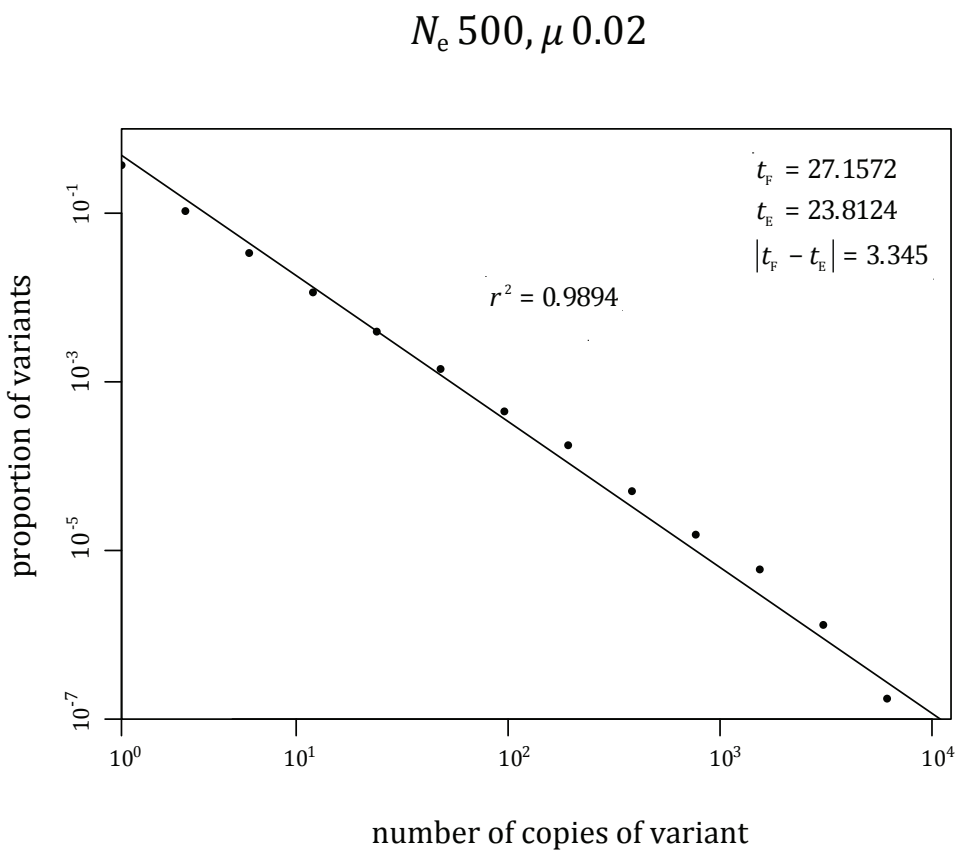
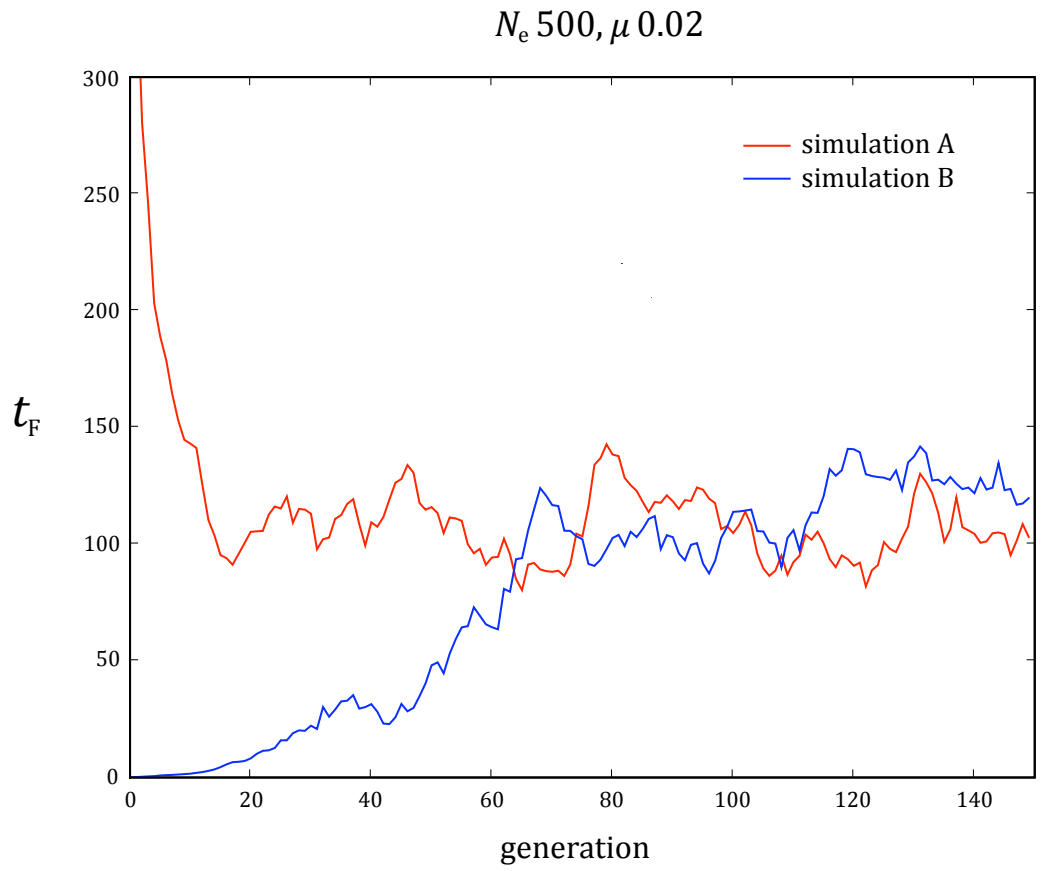


**Figure 4.2:** The empirical heterogeneity statistic $t_F$ over time, for an example pair of neutral model simulations A and B.



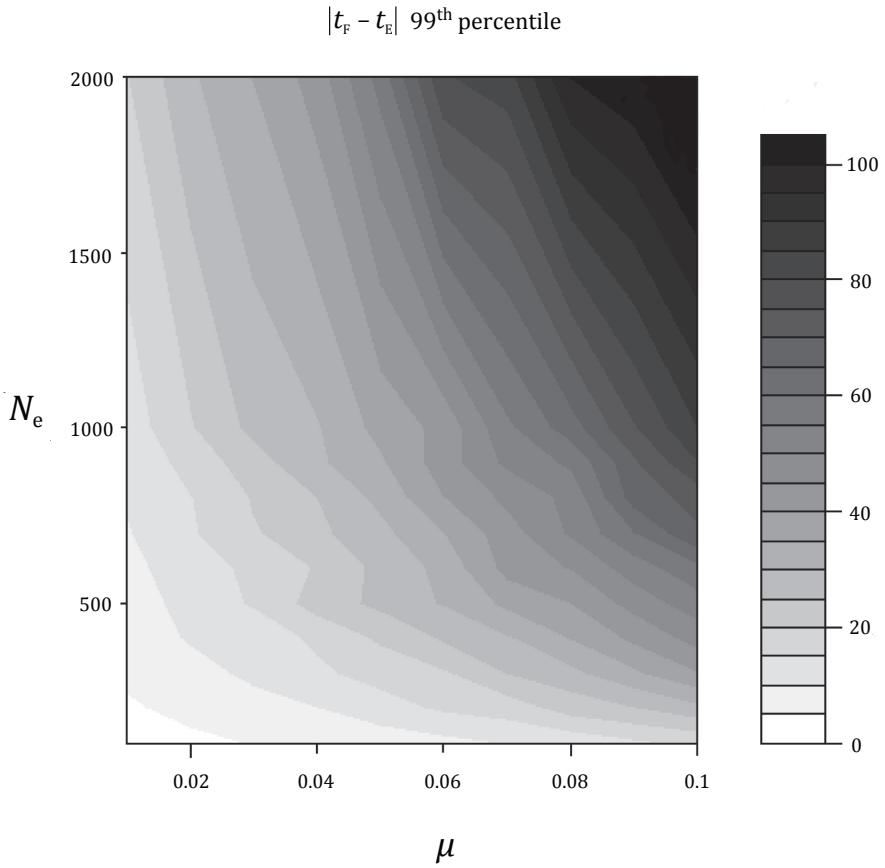

This model also included a parameter that determined the number of previous generations the current generation of copiers had access to i.e. they would sample from the variant frequencies accumulated over those generations. This was intended to provide for the relaxation of one major assumption of previous cultural drift models that is imported from genetics – that only the previous generation is of any importance in the production of the present generation. Cultural transmission has different dynamics to genetic transmission, and accumulated variation from

multiple previous generations is likely to play an important role in cultural transmission processes (Boyd and Richerson 1985, Henrich and McElreath 2003, Tomasello 1999). However, in order for these results to be consistent with those of other models published up to that point I fixed this parameter at 1 throughout, i.e. only the single previous generation was available to copy. Recent work by Bentley *et al.* (2009) has considered this additional parameter – which they term the 'memory' parameter $m$ – in more depth, and shown that the time-depth of variant frequencies that randomly copying agents have access to can affect the types of right-skewed long-tail distribution that arise, with the power-law being simply a special case (Evans 2007). This augmented neutral model can potentially be fitted to a wide range of such distributions found empirically in many socio-economic or -cultural contexts where such a time-frame would be expected to vary (see Bentley *et al.* 2009; section 5).

These neutral simulations showed that the difference $|t_F - t_E|$ increases rapidly as effective population size $N_e$ and innovation rate $\mu$ increase. As an example, the 99$^{th}$ percentile of this absolute difference between the two diversity estimates is shown in figure 4.3 below – giving the criteria for rejecting the neutral model at the 1% level for a large range of population sizes and innovation rates. This is essentially a simulated equivalent to the Ewens-Watterson homozygosity test, which is discussed in the Conclusion section (see also Steele *et al.* 2010). Another notable result of these simulations is that, contrary to previous suggestions, this statistic *does not* appear to be simply a function of $\theta$ (= $4N_e\mu$), and is dependent on the parameter values $N_e$ and $\mu$ *independently* rather than just their product. For example, while $N_e\mu = 50$ in both cases, the 99$^{th}$ percentile for $|t_F - t_E|$ is: 52.76 for $N_e = 500$, $\mu = 0.1$, and 36.77 for $N_e$ =

1000, $\mu$ = 0.05. This effect is far more pronounced when $N_e\mu$ is large, though it still exists at the lowest end of the range, and is most likely due to the much higher (potentially by orders of magnitude) innovation rates that may be expected to operate in cultural contexts compared to the mutation rates of genetics. This simulated rejection criteria may thus be, in itself, an important null result for the application of such neutral models to cultural systems.

**Figure 4.3:** Rejection criteria for the difference between diversity statistics $|t_F - t_E|$, produced through 1,000 iterations of the neutral model, for a range of effective population sizes $N_e$ and innovation rates $\mu$.



### 4.5.3 Application to ancient figure-painted Greek pottery

Schauer (2008) attempts to use both the two estimators of diversity $t_F$ and $t_E$ *and* the full distribution of variant frequencies to determine whether the neutral cultural model best describes the distribution of stylistic variants within a very large archive of ancient figure-painted Greek pottery. This collection, the Beazley archive (see for example Beazley 1956), comprises around 75 thousand figure-painted pots and fragments dated between 650 BC to 300 BC from the *Kerameikos* in Athens, a small district containing the inter-related pottery and pot-painting workshops over this whole period. An ongoing project to computerize this archive was begun in 1979, but appears ill-designed for queries on the large-scale or 'population level' demanded for analysis of this kind. Following the considerable achievement of turning this archive into a searchable database (Schauer 2008; chapter 3), the final dataset comprised around 39,000 useable entries and was split into 10 shape sub-categories (see Schauer 2008; page 66).

Schauer (2008) was able to use the simulated null distributions I generated for the statistics $|t_F - t_E|$ and $r^2$ to reject the neutral model for some specific subsets of his dataset; while for others the neutral model provided an adequate explanation of the stylistic variation observed (see Schauer 2008; chapter 4). In some cases the diversity measure difference $|t_F - t_E|$ and the power-law fit gave contradictory results (e.g. Schauer 2008; page 134); in that the diversity indices calculated from the data suggested rejection of neutrality, while the distribution of variant frequencies fit sufficiently closely to a power-law (as measured by $r^2$) that neutrality could not be rejected. In these cases Schauer (Schauer 2008; page 137) decided to use the more sensitive test for rejecting the neutral model, and in most of these conflicting cases

this was provided by the diversity measure difference $|t_F - t_E|$ being significantly larger than that expected under neutrality.

In conclusion, this work was a novel attempt to reanalyze a well-known Classical dataset with tools developed within evolutionary archaeology and cultural evolution. It successfully demonstrated that the observed diversity of cultural artefacts *can* sometimes be consistent with a neutral drift-like process, and that these processes should, in the right contexts, be taken seriously. However, my subsequent work, described in the next section, has shown that introducing surprisingly high levels of *non-random* copying into an otherwise neutral cultural process can still generate variant frequency distributions that are indistinguishable from those produced under complete neutrality, i.e. the 'neutral' null model might not be so neutral after all.

## 4.6 Frequency-dependent copying biases to the neutral model

### 4.6.1 The model of Mesoudi and Lycett (2009)

So far I have only considered models operating under the assumption of random-copying. While this model may appear to closely describe many real cultural datasets at the population level (e.g. Bentley *et al.* 2004), there is a wealth of accumulating evidence of our propensity to obtain cultural information in a selective way, subject to various evolved psychological biases (see Chapter 1; section 1.2). It is important just to reiterate here that in the neutral model individuals copy the cultural variant of a *randomly* chosen individual, so that variants are adopted through direct

frequency-dependence. This contrasts with frequency-dependent *copying biases*, which are defined by the *non-random* choice of cultural variant; with the most prevalent being preferentially adopted under conformity, or the most scarce under anti-conformity.

In a recent study Mesoudi and Lycett (2009) – hereafter referred to as M&L – explored the effect of incrementally adding a conformity or anti-conformity bias to the neutral model. They found that a conformity bias skewed the variant frequency distribution towards the 'winner take all' distribution. Shennan and Bentley (2003) had expected this same distribution under their model of 'independent decisions', where each variant has an equal probability of being copied regardless of its frequency. In M&L's anti-conformity model, they also generated the distribution one would expect – an increase in variants with mid-range frequencies and a reduction in high-frequency variants i.e. at the far right of the distribution (Mesoudi and Lycett 2009; figure 2B).

M&L also introduced another type of frequency-dependent bias, in which individuals copy at random as in the neutral model, but selectively ignore potential traits that were either in the top *T* fraction ('common-trait trimming') or in the bottom *B* fraction ('rare-trait trimming'). They demonstrated that this 'frequency-dependent-trimming' could result in frequency distributions that appear to be neutrally generated (Mesoudi and Lycett 2009; figure 3A-B); this is examined in more detail below. The 'rare-trait trimming' model did, however, generate a slight increase in high-frequency variants. The probable reason that the biased elimination of low-frequency variants can boost the highest-frequency variants is that it is the low-

frequency variants that eventually push them out through turnover (Bentley *et al.* 2007), so without the eventual competition the current 'winners' can remain longer.

### 4.6.2 Testing for deviation from the neutral 'null' distribution

However, while this work provided an exploration of the effects of copying biases, M&L did not use a formal testing procedure to determine whether the distributions generated by any of these processes deviated significantly from those expected under the neutral model. Applying the hypothesis-testing approach described in section 4.5.2, and utilized by Schauer (2008), to the copying bias models of M&L would allow a more rigorous examination of the effects of frequency-dependent copying on variant frequencies. This approach would make it possible to determine the upper bound for the degree of conformity or anti-conformity bias that could occur within an otherwise randomly-copying population but still generate variant frequency distributions that were statistically indistinguishable from those produced by the neutral model. In essence, this would provide an answer to the rhetorical question posed by Frank: *"What range of non-neutral generative models attract to the common patterns* [i.e. power-law distributions]*, because the extra non-neutral information gets washed out with aggregation?"* (Frank 2009; page 1584, my addition in brackets). Consequently, this approach may provide an important note of caution in interpreting too strongly the fit to a power-law of real-world variant frequency data as diagnostic of a neutral cultural evolutionary process.

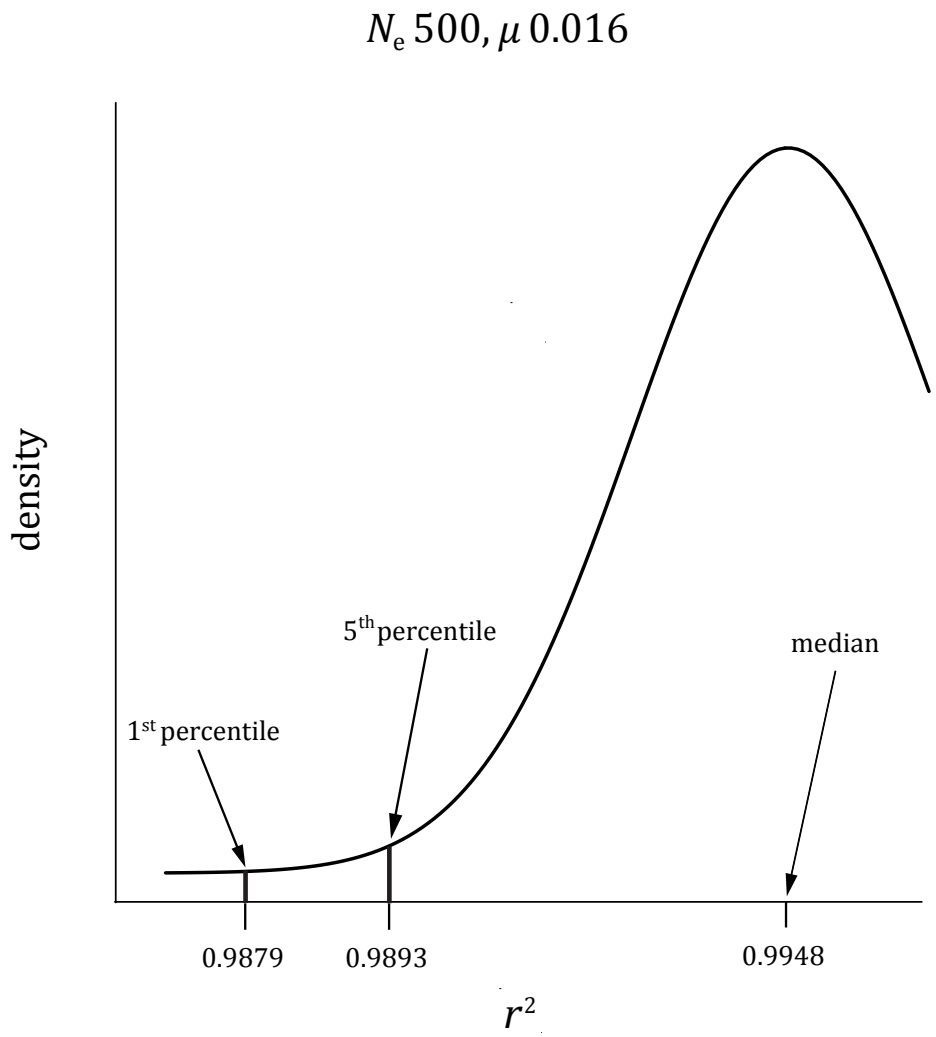
Exactly as before, I first simulated the neutral model 5,000 times for each pair of parameter values presented previously (Bentley *et al.* 2004; table 1) – effective population size $N_e$ and innovation rate $\mu$. The previous related work with this model (section 4.5.2, and Schauer 2008) demonstrated that over this range of parameters equilibrium would be reached in around 300 time steps, so in order to save computational time I simulated just 500 time steps instead of the 1,000 used by previous authors (Bentley *et al.* 2004, Mesoudi and Lycett 2009).

In each iteration I fitted the resulting cultural variant frequencies – under logarithmic binning as described in section 4.5.2 – to a power-law distribution, and through this process generated a simulated distribution of $r^2$ values for the power-law fit (see figure 4.4). This is the distribution we would *expect* given the null model of neutral cultural evolution, i.e. with no bias. I also similarly created the null distribution of the power-law exponent (or slope) $\alpha$, to provide an additional means of rejecting neutrality. However, all subsequent analyses showed that the exponent was far less sensitive to deviations from neutrality than $r^2$, so I used solely $r^2$ as the basis for all subsequent hypothesis tests. Then, following M&L (Mesoudi and Lycett 2009), I created two further simulation models each of which introduce a different biased copying process.

The first model allowed individuals to 'conform' with probability $c_p$ by copying the most frequent variant in the population ('positive frequency-dependence'). The second allowed individuals to pursue anti-conformity with probability $c_n$ by copying the least frequent variant ('negative frequency-dependence'). As in M&L (Mesoudi and Lycett 2009) the remaining individuals copied

at random as previously described. If there were multiple different variants at the highest or lowest frequency then one was chosen at random for the biased learners to copy.

**Figure 4.4:** An idealized plot of the empirical $r^2$ null distribution created from 5,000 neutral simulations, as described in the text.



For each of these two biased-copying models I performed a set of simulations over a range of values for $c_p$ or $c_n$, and for the same parameter values $N_e$ and $\mu$ used previously with the neutral model. For each parameter set ($N_e$, $\mu$, $c_{p/n}$) I performed 500 iterations – in each one fitting a power-law to the resulting variant frequency distribution – and recorded the mean $r^2$ value of these 500 fits (see figures 4.5 and 4.6 for an example of a single iteration). I was then able test for any significant divergence from the null neutral model with increasing levels of biased copying, $c_p$ or $c_n$, using a one-tailed hypothesis test for $r^2$. If the mean $r^2$ value calculated from simulating with

**Figure 4.5:** An example of the variant frequencies generated in one iteration of the model with conformist bias probability $c_p$ = 0.1, and the resulting power-law fit.

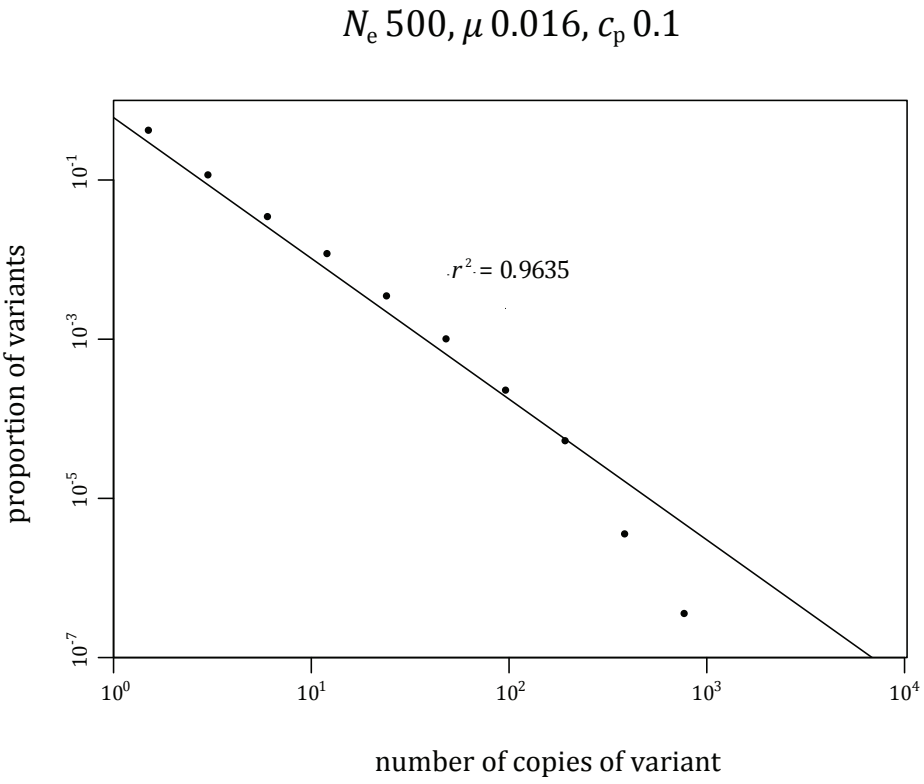
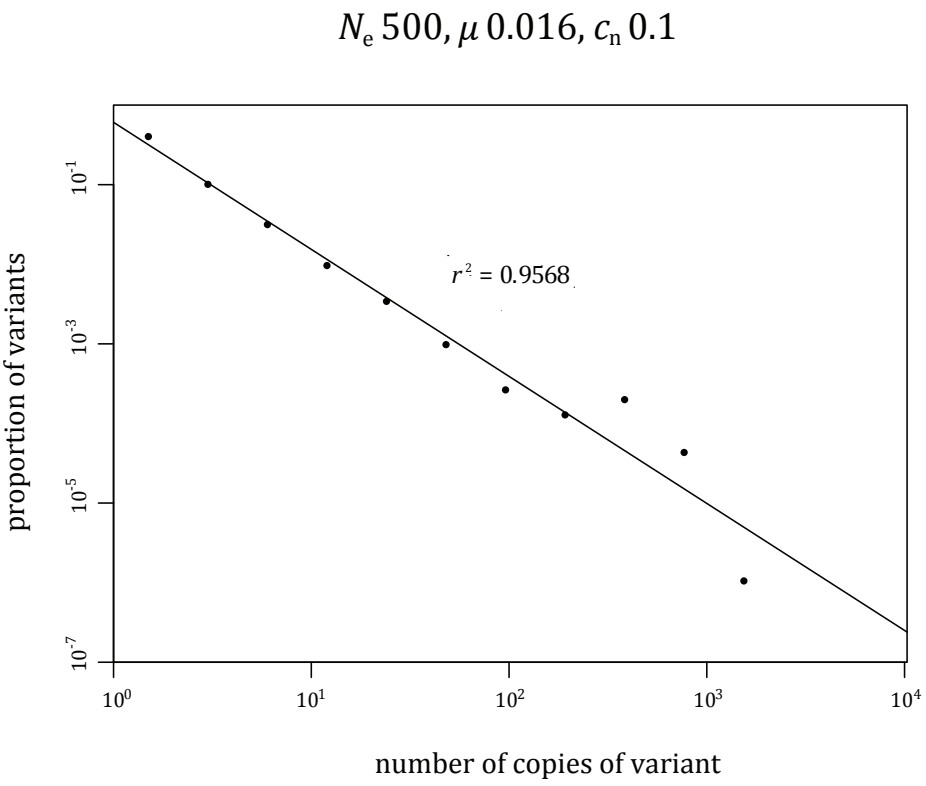


**Figure 4.6:** An example of the variant frequencies generated in one iteration of the model with anti-conformist bias probability $c_n = 0.1$, and the resulting power-law fit.




($N_e$, $\mu$, $c_{p/n}$) was less than the $m^{th}$ percentile of the expected null distribution for ($N_e$, $\mu$) I could reject neutrality at the $m$% level (see figure 4.4). The test is one-tailed because no process I modelled could feasibly *increase* the degree of fit to a power-law; indeed both biases have been repeatedly shown to *reduce* the fit (Mesoudi and Lycett 2009, Bentley and Shennan 2003, Shennan and Wilkinson 2001). As I increased $c_p$ or $c_n$ from zero, the values at which I could reject neutrality gives the upper bound for the level of conformity or anti-conformity bias that could potentially be acting in the population without detecting significant deviation from neutrality.

### 4.6.3 Results and discussion

As figures 4.5 and 4.6 show, and in accordance with M&L (Mesoudi and Lycett 2009), in both the conformist and anti-conformist cases the effect of the bias was to make the variant frequency distribution fit less well to a power-law. Both of the example variant frequency distributions generated with copying bias probabilities of 0.1, given in figures 4.5 and 4.6, could be rejected at the 1% level as being generated by the neutral model. The 1st percentile of the simulated neutral distribution of $r^2$ with $N_e$ = 500 and $\mu$ = 0.016 was 0.9879 (see figure 4.4), which is larger than those of both of these *single iteration* examples, meaning they could comfortably be rejected as being generated from a neutral process.

There is a crucial difference, however, in the sensitivity of the distribution to these two biases. Anti-conformity bias causes the power-law to degrade much faster than does conformity; increasing those variants in the low- or mid- frequency range and decreasing those at the upper end, resulting in the 'humped' distribution (figure 4.6) which is also clearly demonstrated in M&L (Mesoudi and Lycett 2009; figure 2B). In fact the distributions that result from surprisingly high levels of conformist bias can sometimes be difficult to distinguish from neutral random-copying. As the model starts at maximum heterogeneity, i.e. all $N_e$ initial variants are equally prevalent at frequency $1/N_e$, conformity bias has the effect of rapidly amplifying any early random changes in variant frequencies caused initially through drift. This provides a selective buffer against loss through drift for any such lucky variant, and simply propels it early into relatively safe territory within a population that is still largely comprised of frequency-dependent random copiers.

Tables 4.1 and 4.2 give the minimum bias levels under which the null neutral model could *first* be rejected; with conformity bias probability $c_p$ and anti-conformity bias probability $c_n$ respectively. These results show that it is possible for a cultural evolutionary process to have quite substantial levels of biased copying while still resulting in a variant frequency distribution indistinguishable from that generated from a neutral random-copying process using the power-law fit as a test statistic. Indeed, under certain parameter sets there can be up to ~15% of individuals learning with conformist bias, which will not be statistically detectable. While anti-conformity bias is more easily detected at lower levels – again in agreement with M&L (Mesoudi and Lycett 2009) – it is surprising that up to 5% of individuals may be copying in this way undetected, under relatively plausible parameter values.

**Table 4.1:** Minimum conformity bias $c_p$ at which neutral random-copying could be rejected at 1%.

| $\mu$ | $N_e = 125$ | $N_e = 250$ | $N_e = 500$ | $N_e = 1000$ |
|---|---|---|---|---|
| 0.004 | 0.1 | 0.06 | 0.04 | 0.03 |
| 0.008 | 0.08 | 0.04 | 0.03 | 0.03 |
| 0.016 | 0.06 | 0.05 | 0.04 | 0.06 |
| 0.032 | 0.08 | 0.11 | 0.06 | 0.08 |
| 0.064 | 0.12 | 0.15 | 0.17 | 0.11 |
| 0.128 | 0.34 | 0.19 | 0.22 | 0.26 |

**Table 4.2:** Minimum anti-conformity bias $c_n$ at which neutral random-copying could be rejected at 1%.

| $\mu$ | $N_e = 125$ | $N_e = 250$ | $N_e = 500$ | $N_e = 1000$ |
|---|---|---|---|---|
| 0.004 | 0.03 | 0.02 | 0.01 | 0.01 |
| 0.008 | 0.02 | 0.01 | 0.01 | 0.01 |
| 0.016 | 0.02 | 0.02 | 0.01 | 0.02 |
| 0.032 | 0.02 | 0.03 | 0.02 | 0.03 |
| 0.064 | 0.03 | 0.05 | 0.16 | 0.12 |
| 0.128 | 0.29 | 0.2 | 0.13 | 0.12 |

In conclusion, these results demonstrate that datasets previously deemed a product of a neutral process might potentially have been generated by processes involving a significant non-random component but that have simply not climbed far enough out of the power-law 'attractor' basin (Frank 2009) to be noticeable. The two tables presented above give an approximate attractor boundary for each parameter combination; above which neutrality can be confidently rejected, and below which the non-random processes will simply be disguised by population-level aggregation.

At the higher innovation rates ($\mu \geq 0.064$) the level of innovation occurring *can* disguise the copying biases to a larger degree, resulting in higher critical minimums for neutrality rejection. However, it should be noted that this effect is very much prone to the stochastic nature of innovation events and these results may in some cases appear anomalous (e.g. anti-conformity $N_e = 500$, $\mu = 0.128$).

A further use of the approach presented here is in its application to frequency distributions generated through the process of 'frequency-dependent trimming'

introduced by M&L. This is where individuals are allowed to copy randomly but selectively ignore potential traits that were either in the top *T* fraction, 'common-trait trimming', or in the bottom *B* fraction, 'rare-trait trimming'. M&L simulated models with $T = 0.4$ and with $B = 0.4$ and demonstrated that both variant frequency distributions thus produced appeared to fit well to a power-law. However, comparing the $r^2$ values of the two fits as presented in M&L (Mesoudi and Lycett 2009; figure 3) to the simulated null $r^2$ distribution allows the formal rejection of the neutral model at the 1% level. The authors use $N_e = 250$ and $\mu = 0.008$ throughout, and the 1st percentile of the expected $r^2$ distribution for these parameter values is 0.990. This allows the rejection of the null model for $B = 0.4$, with $r^2 = 0.975$, and for $T = 0.4$, with $r^2 = 0.989$. So even though frequency-dependent trait trimming appears to generate seemingly neutral-like variant distributions, my approach enables the identification of a non-random / selective process operating in the population.

## 4.7 Conclusion

While previous studies have attempted to use the diversity statistics $t_F$ and $t_E$ or the fit to a power-law as evidence for or against neutrality, none had so far addressed the need to test for significance of these statistics against the distribution expected under neutrality over a range of population sizes and innovation rates. I recreated the neutral model of Bentley *et al.* (2004) in order to generate such null distributions for a wide range of parameter values (section 4.5), and these were used by Schauer (2008) to test for significant deviations from neutrality in subsets of a large decorated pottery dataset.

The next section (4.6) considered the effect of introducing parameterized levels of frequency-dependent copying into an otherwise neutral model. My results showed that the neutral model may not be such a simple null model after all, as it appears that population-level aggregation can disguise quite significant levels of *non-random* biased copying by generating variant frequency distributions indistinguishable from those of the neutral model (see also Frank 2009). This work shows that only when these copying biases reach a critical threshold in the population can complete neutrality be *rejected* as the underlying generative model, and thus provides a note of caution in over-interpreting a dataset as neutrally generated.

The neutral model I have utilized here, based on the work of Neiman (1995) and Bentley and colleagues (e.g. Bentley *et al.* 2004), retains two major assumptions carried over from the population genetic model from which it is derived; namely, that cultural transmission proceeds in discrete generations, and that an equilibrium between cultural drift and innovation is achieved. While both assumptions can in many contexts appear reasonable, the model developed in the following chapter (Chapter 5) demonstrates that neither is *necessary* to construct a tractable cultural evolutionary model that can be used to estimate population parameters from archaeological data.

A paper published in 2010 (Steele *et al.* 2010) covered some of the same ground as section 4.5 of this chapter. Steele *et al.* noted that the Ewens-Watterson homozygosity test (Watterson 1977, 1978, see also Slatkin 1994, 1996) provides a way of determining whether the empirical homogeneity statistic $\hat{F}$ is significantly different from the homogeneity expected under neutrality. This constitutes a more

formal approach than that of Shennan and Wilkinson (2001), and is directly equivalent to testing for significance of the difference $|t_F - t_E|$. Had I been aware of the Ewens-Watterson test when this work was started, in 2007, it may have served as the basis of my approach. While the work presented in section 4.5.2 is an empirically generated equivalent to this test, the fact that the test statistic is dependent on the actual values of $N_e$ and $\mu$ rather than just their product – as assumed by the Ewens-Watterson test – is an important additional result.

Steele *et al.* used the Ewens-Watterson test to analyze changing bowl rim shapes in a large Hittite ceramic dataset (~1,800–1,200 BC) from central Anatolia (Steele *et al.* 2010; page 4). They found that while they were unable to reject neutrality with this variant frequency distribution approach, a strong correlation between the frequency of rim variants over time and the ware type used *did* suggest some functional, or 'selective', change. They concluded that, in line with the genetics literature (see Steele *et al.* 2010; page 3), detecting departure from neutrality might require more statistically powerful tests, and further provide a note of caution against uncritically assuming neutrality of a typological trait.

# Chapter 5

# A cultural drift model: and an approximate maximum likelihood estimation algorithm

## 5.1 Summary

This chapter provides the details of a novel cultural drift model – based on the Moran model of population genetics – that relaxes two important assumptions made by many existing cultural evolutionary models. The model avoids the issue of discrete 'cultural generations' by introducing the idea of cultural half-life, and also avoids a reliance on the assumption of cultural equilibrium. Given a minimal archaeological dataset this model, under the assumption of neutral cultural drift, allows the use of an approximate maximum likelihood algorithm to accurately estimate two important parameters: the effective size of the population of cultural artefacts and the half-life of this population. Inferring these parameters may provide information on the cultural dynamics of past human populations. It is successfully tested with simulated datasets designed to replicate observed data, and then applied to a real archaeological dataset of decorated LBK vessels from southwest Germany.

## 5.2 Abstract

In this chapter I develop a model that allows approximate maximum likelihood estimation of two parameters of interest from a small, observed archaeological dataset. The model uses Monte Carlo simulation to estimate the joint density $f(\kappa, \tau)$, where $\kappa$ is the maximum frequency a cultural variant reaches and $\tau$ is the time (in years) to this maximum from first appearance, within a pool of cultural artefacts of effective size $N_e$ under the assumptions of cultural drift. Simulation time-steps are scaled into time in years through the additional half-life parameter, $t_{1/2}$. These joint densities are estimated for a wide range of parameter combinations ($N_e$, $t_{1/2}$), where $10 \leq N_e \leq 10{,}000$ and $1 \leq t_{1/2} \leq 199$.

Then, given a small sample of $s$ observed values in the form ($\kappa_i$, $\tau_i$) calculated from archaeological assemblage data, likelihood values are computed for each combination of parameter values ($N_e$, $t_{1/2}$). By comparing these likelihoods over this grid of parameter values an approximate maximum likelihood estimate (MLE) $(\hat{N}_e, \hat{t}_{1/2})$ is obtained for the unknown parameters given this observed data. This basic algorithm is extended to incorporate innovation, at rate $\mu_{year}$ per year. Though the estimation of parameter values $(\hat{N}_e, \hat{t}_{1/2})$ can be improved with an empirical estimate of $\mu_{year}$ from the data, this is *not necessary* if the estimated rate is ≤2%. Results for simulated test observed data sets are given to demonstrate the efficacy of the method, before its application to a real archaeological dataset from the LBK of southwest Germany.

## 5.3 Introduction

There are two major assumptions common to the neutral cultural evolutionary models reviewed in the previous chapter (Chapter 4; section 4.3): firstly, that innovation / drift equilibrium of cultural variant frequencies is reached (e.g. Neiman 1995, Bentley *et al.* 2004, Kandler and Laland 2009, Strimling *et al.* 2009); and secondly, that cultural systems evolve in a series of discrete non-overlapping generations (e.g. Neiman 1995). These two are, of course, inextricably related. If 'cultural generations' are long-lasting, a cultural system would have to endure for a long time – under the same conditions – in order to allow a sufficient number of generations to pass in order for equilibrium to be approached. Both of these assumptions are carried over from population genetics; where the ubiquitous Wright-Fisher model (Fisher 1930, Wright 1931, also see Ewens 2004 for a review) insists on non-overlapping generational replacement; and where equilibrium is generally invoked in order to make predictions about the long-term fate of an allele. However, aside from modeling convenience there does not appear to be much justification for the notion of the discrete generational replacement of cultural artefacts. Similarly while there is plenty of evidence that some cultural evolutionary processes do operate on a shorter time-scale than those of genetics (see for example Boyd and Richerson 1985, Richerson and Boyd 2005), it is still not clear whether equilibrium of stylistic cultural variants is common or that it should be considered an inevitable cultural evolutionary outcome. Both of these assumptions may be *reasonable* in terms of constructing simplified – and even analytical – models, but the following model demonstrates that they may be *unnecessary* in providing good inferences on cultural evolutionary processes.

The model developed in this chapter is intended to relax these two common assumptions, and to provide for adequate estimation of the parameters of the idealized underlying cultural evolutionary process from minimal archaeological data. The parameters we are interested in are the idealized, or 'effective', *size* of the population of cultural artefacts $N_e$ and the *half-life* of this pool $t_{1/2}$ that serves as a tractable proxy for the turnover rate in the population. The data required for this inference process comprises of the maximum sample frequency and the time from first appearance to this maximum, i.e. the battleship curve 'peaks', for a small number of different variants within a single assemblage. An empirical estimate of innovation rate, while not necessary, helps improve the parameter estimation and is only required at all if this estimated rate is >2% per year.

Given this data, parameter inference relies on the assumption of cultural trait neutrality – more of which below and in Chapter 4. Conversely, this model can use this data in conjunction with empirically estimated (or constrained) population size and half-life parameter values in order to either accept or reject cultural neutrality. As far as I am aware, there is no pre-existing model or analysis technique that attempts to use minimal archaeological seriation data in this way.

## 5.4 The cultural drift model

### 5.4.1 Motivation for the model

The Wright-Fisher model (Fisher 1930, Wright 1931, also see Ewens 2004) is the most well-known, and perhaps the most commonly utilized, model in population genetics. As a result, many cultural evolutionary models use it as a basis for derivation (for example see Neiman 1995, Bentley *et al.* 2004) and carry over some assumptions that may not be necessarily applicable to modeling the evolution of culture. One such basic assumption of the Wright-Fisher model is that time proceeds in discrete generational steps, and that in each of these generations the entire population is replaced by an offspring generation. While this idealized model is more than adequate for describing the genetic evolution of a population, its insistence on discrete generational replacement makes it potentially problematic for direct application to the evolution of culture, as the notion of a cultural generation is not nearly so obviously defined as those of biology (see for example Boyd and Richerson 1985). Another population genetic model, that of Moran (1958, 1962, also see Ewens 2004), may be more appropriate in this regard as it does not insist on full population replacement at any point, and has a less strictly defined notion of a generation. Here, I describe how this model can be adapted for modeling cultural evolution.

### 5.4.2 The two-allele Moran model of population genetics

The very simplest version of the Moran model (Moran 1958, 1962) concerns a population of haploid individuals of effective size $N_e$. In each time-step $t = 0, 1, 2, \ldots$ an

individual is randomly chosen to reproduce, which in this haploid case means simple duplication. Then one individual, which can be the reproducing one, is chosen to die and is removed from the population. If we assume this population is comprised of individuals carrying one of two possible alleles $A_1$ or $A_2$, then in any one time period there are only three possibilities: allele $A_1$ can increase in count by 1, allele $A_2$ can increase in count by 1, or both can stay the same. Restricted to just these two possible alleles, any increase in one allele can be expressed simply as a decrease in the other, so we can characterize the behaviour of the entire population through time using the count (or frequency) of one arbitrarily chosen focal allele. Changes in the count $c$ of this focal allele in the population is modeled through time, with the three mutually exclusive possibilities in each time-step determined by the transition probabilities: $p_{c,\,c+1}$, $p_{c,\,c-1}$, $p_{c,\,c}$ . These probabilities are not described here as near-identical versions for our cultural model are given below.

In each time-step an individual is replaced with probability $Pr = 1/N_e$ so the lifespan of an individual is geometrically distributed with mean = $1/Pr = 1/(1/N_e) = N_e$ time-steps. These $N_e$ time-steps could comfortably be interpreted as an overlapping 'generation' in a biological context, but from a culture-evolutionary perspective each individual time-step can be considered as the destruction or discarding of a cultural artefact and the creation of its replacement within a population, or pool, of artefacts. We assume that this pool is comprised of cultural artefacts differentiated solely by a neutral trait e.g. a pool of vessels each of which bears a stylistic design element that may vary. The frequencies of these various stylistic types, or 'cultural variants', can be modeled in an analogous way to the alleles of a population genetic model (see for example Boyd and Richerson 1985, Cavalli-

Sforza and Feldman 1981, Neiman 1995). Each time-step can be considered as a 'copying event' where the variant type that is created as a replacement is determined by a random copying, or cultural drift, process. Thus these discrete time-steps / copying events are themselves the appropriate time unit, minimizing the problem of 'cultural generations'. By this rationale, the Moran model is a better starting point than any existing alternative.

### 5.4.3 The basic cultural drift model

#### 5.4.3.1 The model and simulation process

We start with a population of cultural artefacts of effective size $N_e$ and introduce a single novel variant into this pool i.e. at frequency $1/N_e$. Similarly to the above, we use $c$ to track the *count* of this focal variant in the population and $p_{x,y}$ for the three transition probabilities at each time step. We assume that the effective population size is fixed at $N_e$ throughout, but unlike the above two-allele Moran model we assume that in each time-step one artefact is removed first *before* replacement occurs, i.e. the population temporarily reduces in size to $N_e - 1$. This is to recreate the intuitive cultural idea that an artefact is created – with type dependent on the variant frequencies of those remaining – in order to replace the one destroyed which, unlike in a genetic model, plays no part in determining its type. Under these assumptions, the (transition) probability that focal variant count $c$ will increase to $c + 1$ in the next time-step is:

$$p_{c,c+1} = \frac{c(N_e - c)}{N_e(N_e - 1)}$$

In the absence of innovation in this basic model, it follows that by symmetry the probability that count $c$ will decrease by one to $c - 1$ is the same:

$$p_{c,c-1} = \frac{c(N_e - c)}{N_e(N_e - 1)}$$

Finally, the probability that count $c$ remains unchanged is:

$$p_{c,c} = \frac{N_e(N_e - 2c - 1) + 2c^2}{N_e(N_e - 1)}$$

$$= \frac{(N_e - c)^2 + c^2 - N_e}{N_e(N_e - 1)}$$

which is equal to 1 minus the sum of the first two probabilities. The superficial differences with the basic two-allele Moran model are simply due to the temporal reordering of the birth / death process, which we deem more appropriate for this cultural model.

Whilst using a formulation that would lend itself well to a Markov chain approach, I was not concerned with long-term equilibrium behaviour but rather the extreme values of the focal variant count ($c_{max}$) and the time-step at which this occurred ($g_{max}$). Thus there was no requirement for equilibrium to be reached; I simply recorded the extreme frequency values reached in each iteration of this dynamic process. So with $c = 1$ (i.e. a novel variant) in the first time-step, I simulated forward with the above transition probabilities, recalculated at each step, for a total

of 6,000 time-steps. I calculated and recorded the maximum frequency κ (= $c_{max}/N_e$) that this focal variant reached and the time-step $g_{max}$ at which this maximum occurred. Repeating this simulation process 1 million times resulted in a large dataset of (κ , $g_{max}$) values for the given population size $N_e$.

#### 5.4.3.2 Half-life and simulation time scaling

In order to relate these simulation time-steps to real time I introduced a half-life parameter $t_{1/2}$ defined in years. By definition this is the time taken for exactly half of the population (i.e. $N_e/2$) to have been removed, but note I replace an artefact at every step to maintain the population at size $N_e$. It should also be noted that the following calculations do not alter in any way the fixing of population size at $N_e$ throughout each simulation, but are simply for calculating the scaling factor $t_{step}$.

The length of a time-step in years $t_{step}$ needs to be calculated in terms of a given half-life $t_{1/2}$ (see figure 5.1). Under the canonical exponential decay model, with decay constant λ, the population size at time $t$ $N_e(t)$ can be expressed in terms of the initial population size $N_e(0)$ as:

$$N_e(t) = N_e(0)\, e^{-\lambda t}$$

Given that at time $t_{1/2}$ an exponentially decaying population would have size $N_e(0)/2$:

$$N_e(t_{1/2}) = N_e(0)\, e^{-\lambda t_{1/2}} = N_e(0)/2$$

So:

$$\lambda t_{1/2} = \ln(2)$$

$$\lambda = \frac{\ln(2)}{t_{1/2}} \qquad (5.1)$$

Then, as one replacement takes place each time-step of a simulation, nominally we go from the initial population size $N_e(0)$ to population size $N_e(0) - 1$. So similarly:

$$N_e(t_{step}) = N_e(0)\, e^{-\lambda t_{step}} = N_e(0)\left(\frac{N_e(0)-1}{N_e(0)}\right)$$

And:

$$e^{-\lambda t_{step}} = (N_e(0)-1)/N_e(0)$$

$$\lambda t_{step} = \ln\left\{N_e(0)/(N_e(0)-1)\right\}$$

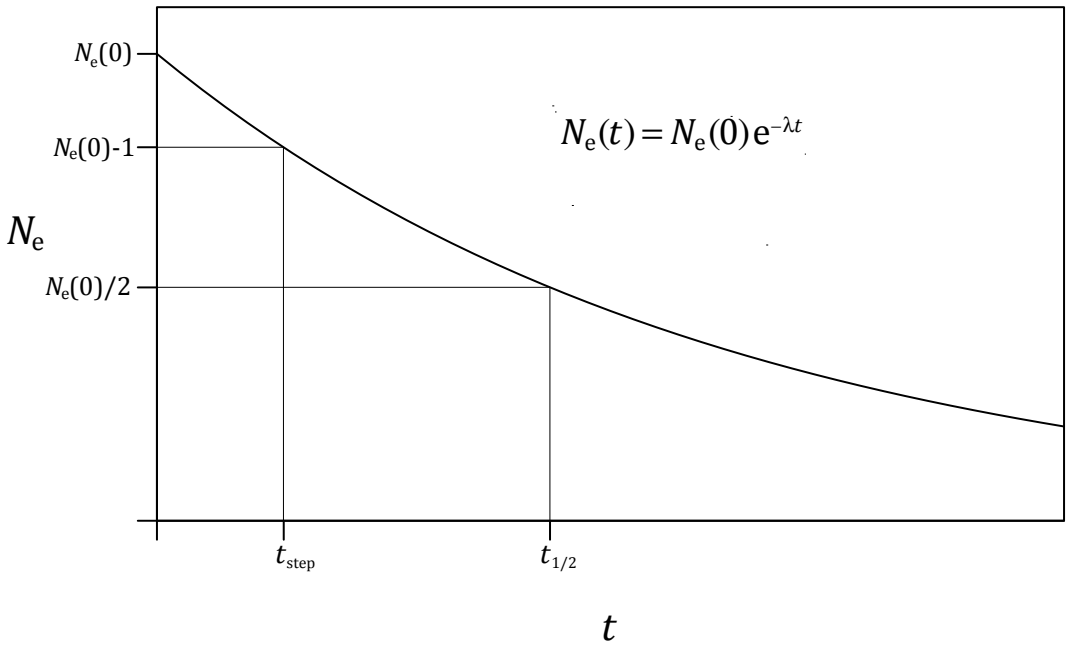
So:

$$t_{step} = \frac{\ln\left\{N_e(0)/(N_e(0)-1)\right\}}{\lambda}$$

$$= \frac{\ln\left\{N_e(0)/(N_e(0)-1)\right\}}{\ln(2)} t_{1/2} \qquad \text{(substituting in equation 5.1)}$$

Using this scaling factor $t_{step}$ for a given half-life $t_{1/2}$, all $g_{max}$ values from the simulated dataset were rescaled into years $\tau$. This rescaled simulated dataset of values ($\kappa_i$ , $\tau_i$) comprised of 1 million independent random draws from the underlying – and, as far as we are aware, unknown – joint probability density surface $f(\kappa , \tau)$.

**Figure 5.1:** The exponential decay model; demonstrating the calculation of the time scaling factor $t_{step}$.

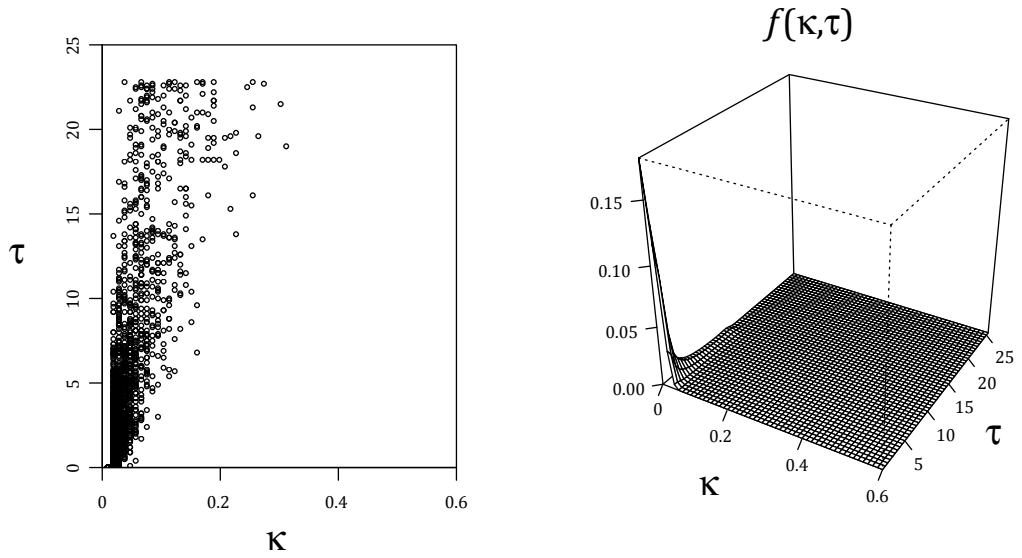



### 5.4.3.3 Estimation of joint density surfaces

These independent draws were used to create the approximate joint density, by performing kernel density smoothing – as implemented in the "KernSmooth"

package (Wand and Ripley 2009) for the R programming language (R Development Core Team 2009, URL: http://www.R-project.org/) – and normalizing so that the density under the approximated surface summed to 1. An example joint density for an arbitrary combination of parameters $N_e$ and $t_{1/2}$ is given in figure 5.2. The time-scale was capped at 500 years, and the result of the kernel density estimation was projected onto a 500 by 500 grid, which extensive preliminary testing suggested was a reasonable balance between the required precision and computational processing time.

**Figure 5.2:** A truncated example of the estimated joint density $f(\kappa, \tau)$ for an arbitrary combination of parameters $N_e$ and $t_{1/2}$. The left-hand plot is a scatterplot of just 5,000 independent draws, whilst the right-hand plot is the joint density $f(\kappa, \tau)$ estimated from the full 1 million simulated. Both x- and y- axes in each plot have been truncated purely for clarity.



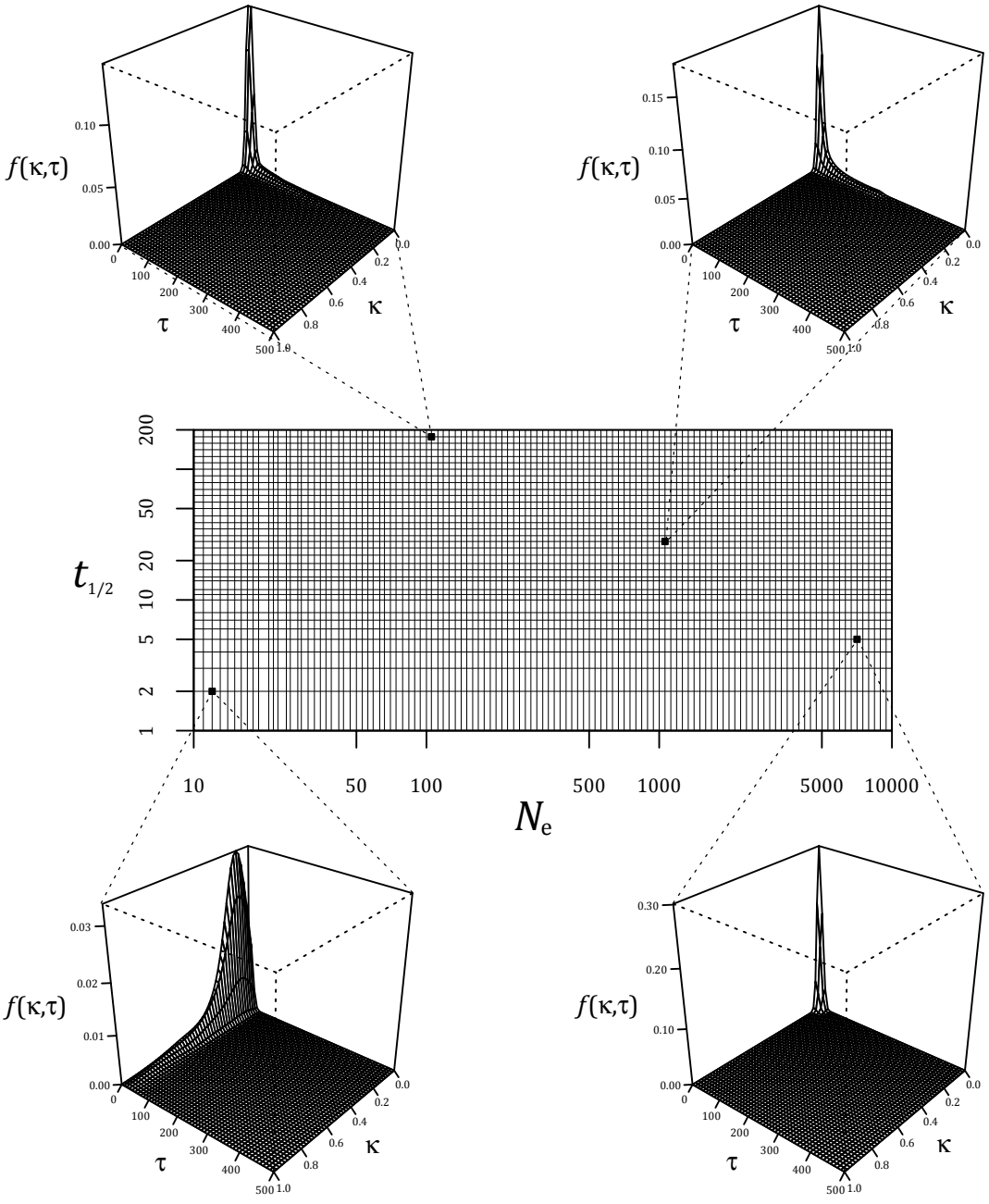
This entire process of simulation and joint density estimation was repeated for a range of values for $N_e$ and $t_{1/2}$ on a grid. I simulated under 118 different effective population sizes in the range [10, 10,000], in combination with 35 different half-life values in the range [1, 199], both on a logarithmic scale. This resulted in a grid of 4,130 joint probability density surfaces $f(\kappa, \tau)$ each estimated from their independently simulated 1 million random draws, as shown in figure 5.3.

Now given an observed archaeological dataset I could use this grid of estimated probability densities to calculate the likelihood of each combination of parameters $N_e$ and $t_{1/2}$, and thus find the parameter values which best explain the data.

#### 5.4.3.4 Likelihood and maximum likelihood estimation

As a brief recap on likelihood: suppose we have a density function $f(D \mid \theta)$, which gives us the probability of observing data $D$ given some values of the parameters $\theta$. Now suppose that we have actually observed this data $D$, then the conditional probability of this data given the parameters $\theta$ is called the 'likelihood', and is this same density function but now considered as a function of the parameters. So the likelihood $L(\theta; D) \propto f(D \mid \theta)$ to some arbitrary constant and as we are *comparing* likelihoods for the same data the constant of proportionality can be safely disregarded. The value of the parameters $\theta$ that maximize this likelihood function is called the joint maximum likelihood estimate (MLE) $\hat{\theta}$.

**Figure 5.3**: The grid of estimated joint density surfaces $f(\kappa, \tau)$, over the full range of values for $N_e$ and $t_{1/2}$. Four examples of different joint densities, taken from the grid where indicated, are given in the subplots – note that the z-axes are on different scales.



#### 5.4.3.5 Likelihood calculation: an example

Now imagine we had a dataset of *s* archaeological observations in the form ($\kappa_i$ , $\tau_i$), where $\kappa_i$ is the maximum observed sample frequency of variant *i* and $\tau_i$ is the length of time in years from its first appearance to this maximum frequency (i.e. random draws from the underlying distribution $f(\kappa , \tau)$). These data are just the peaks and times from minimum frequency to the peak of the 'battleship curves' for *s* different variants within a single assemblage. Given this data, the likelihood of a specific combination of parameters $L(N_e , t_{1/2})$ can be calculated. As the likelihood values are usually extremely small I used log-likelihoods, $\ell(N_e, t_{1/2}) = \ln\left\{L(N_e, t_{1/2})\right\}$, which are calculated as follows:

$$L(N_e, t_{1/2}; \kappa_1, \tau_1, ..., \kappa_s, \tau_s) = f(\kappa_1, \tau_1, ..., \kappa_s, \tau_s \mid N_e, t_{1/2})$$

$$= \prod_{i=1}^{s} f(\kappa_i, \tau_i \mid N_e, t_{1/2})$$

$$\ell(N_e, t_{1/2}; \kappa_1, \tau_1, ..., \kappa_s, \tau_s) = \sum_{i=1}^{s} \ln\left\{f(\kappa_i, \tau_i \mid N_e, t_{1/2})\right\}$$

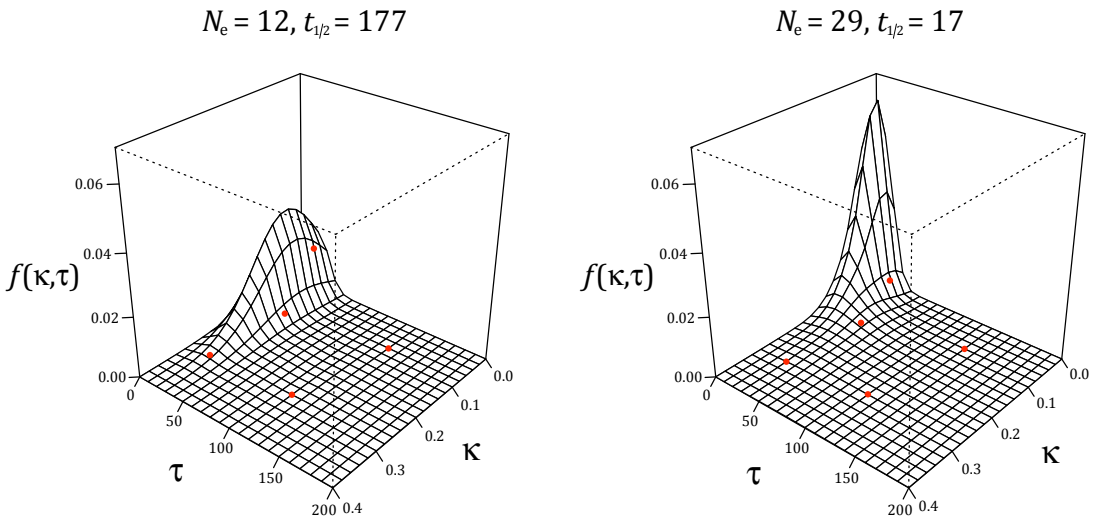
As a concrete example, imagine that this dataset consisted of 5 dummy observations ($\kappa_1$ , $\tau_1$),($\kappa_2$ , $\tau_2$),..., ($\kappa_5$ , $\tau_5$), with values given in table 5.1. For the sake of clarity, further imagine that I have joint density surfaces $f(\kappa , \tau)$ for just two combinations of $N_e$ and $t_{1/2}$: $N_e = 12$, $t_{1/2} = 177$ and $N_e = 29$, $t_{1/2} = 17$ (see figure 5.4).

**Table 5.1:** An example of the calculation of (log-)likelihoods for two parameter sets given five dummy observations ($\kappa_i$ , $\tau_i$).

| | | | $f(\kappa_i, \tau_i)$ | | $\ln\{f(\kappa_i, \tau_i)\}$ | |
|---|---|---|---|---|---|---|
| sample | $\kappa_i$ | $\tau_i$ | $N_e = 12$<br>$t_{1/2} = 177$ | $N_e = 29$<br>$t_{1/2} = 17$ | $N_e = 12$<br>$t_{1/2} = 177$ | $N_e = 29$<br>$t_{1/2} = 17$ |
| $s_1$ | 0.30 | 30 | 0.0028 | 0.0005 | -5.8874 | -7.6736 |
| $s_2$ | 0.16 | 50 | 0.0076 | 0.0044 | -4.8772 | -5.4317 |
| $s_3$ | 0.06 | 40 | 0.0209 | 0.0096 | -3.8689 | -4.6468 |
| $s_4$ | 0.08 | 130 | 0.0002 | 0.0000* | -8.4964 | -10.6780 |
| $s_5$ | 0.28 | 110 | 0.0004 | 0.0005 | -7.8978 | -7.5568 |
| | | | $\ell(N_e, t_{1/2}) = \sum_{i=1}^{s} \ln\{f(\kappa_i, \tau_i)\}$ | | -31.0277 | -35.9869 |

The calculation of log-likelihoods $\ell(N_e, t_{1/2})$ for each parameter combination is given in table 5.1. The asterisked (*) cell in table 5.1 for $f(\kappa_4 , \tau_4)$ is not actually zero – the value was simply rounded down – but the approximated density $f(\kappa , \tau)$ can, in places, be truly zero rather than just very small. In these cases, to avoid problems with taking logarithms of zero, $f(\kappa_i , \tau_i)$ was set at an arbitrarily very small value, calculated as the minimum non-zero value in $f(\kappa , \tau)$ divided by 1,000.

**Figure 5.4:** Estimated joint density surfaces $f(\kappa, \tau)$ for the two example parameter combinations, truncated on both x- and y- axes purely for clarity. The five dummy observations $(\kappa_i, \tau_i)$ given in table 5.1 are marked as red points.




So given this dummy dataset $\ell(N_e = 12, t_{1/2} = 177)$ = -31.0277 and $\ell(N_e = 29, t_{1/2} = 17)$ = -35.9869, and we could conclude that – given this artificial restriction to just two possible parameter sets – that $N_e$ = 12 and $t_{1/2}$ = 177 would be the most likely parameter values given the data.

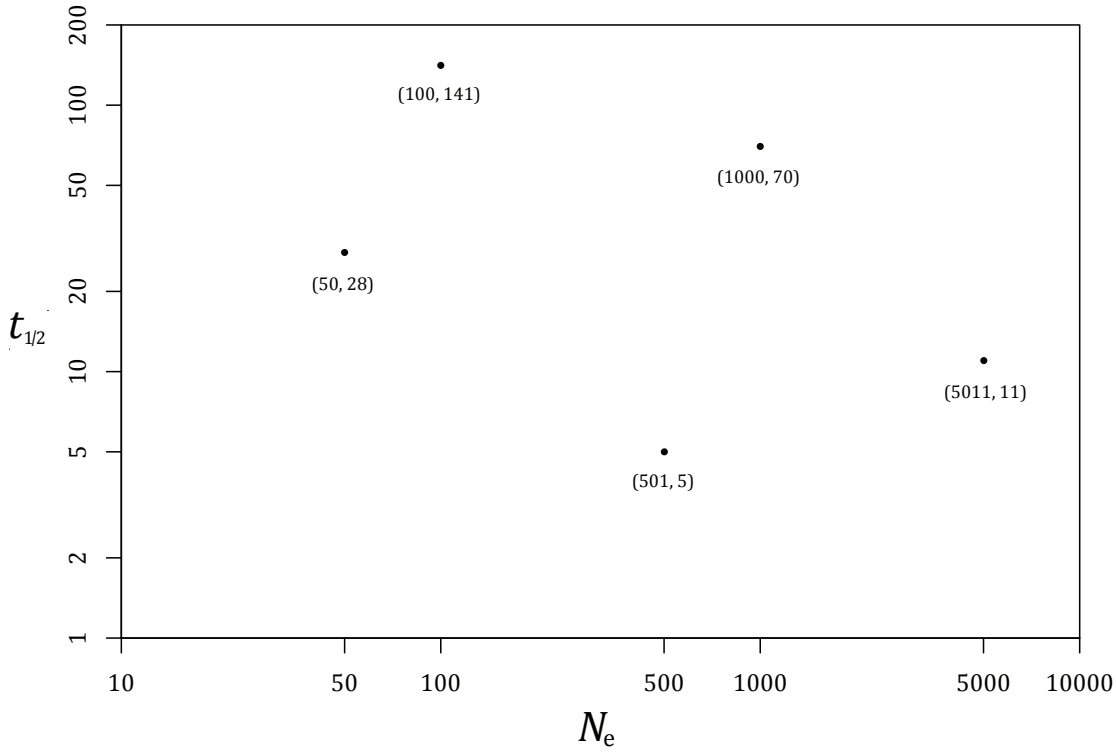
This is just a simple illustrative example, and using the full set of 4,130 joint densities $f(\kappa, \tau)$ I have estimated would allow the calculation of an approximate maximum likelihood estimate (MLE) $(\hat{N}_e, \hat{t}_{1/2})$ for the true parameter values as:

$$(\hat{N}_e, \hat{t}_{1/2}) = \text{argmax}\, \ell(N_e, t_{1/2}; \kappa_1, \tau_1, \ldots, \kappa_s, \tau_s)$$

#### 5.4.3.6 Simulated test datasets

In order to demonstrate this inference method more broadly, I present the results from five simulated test datasets to show that the true parameters ($N_e$, $t_{1/2}$) used to create these dummy 'observed' datasets could be inferred with great accuracy. The five dummy parameter sets are reasonably representative of the whole parameter space (see figure 5.5). Although the method appears robust across the entire parameter space, the processing time involved has restricted detailed exploration to these five examples.

**Figure 5.5:** The five test parameter sets represented in the full parameter space for $N_e$ and $t_{1/2}$.



#### 5.4.3.7 Results for simulated test datasets

**Test dataset 1: $N_e$ = 100, $t_{1/2}$ = 141**

I first simulated an independent set of 100 'observations' of the form ($\kappa_i$ , $\tau_i$) with parameter values $N_e$ = 100, $t_{1/2}$ = 141. This sample size $s$ of 100 was chosen purely for demonstration, and the effect of sample size on the quality of inference will be discussed in more detail below. It did however appear that under this basic model a very small sample ($s$ = 10) was sufficient to accurately recapture the simulated parameters. Given this simulated data set, I then used the method as described, but now using the full set of joint densities $f(\kappa , \tau)$, to calculate log-likelihoods for each ($N_e$ , $t_{1/2}$) combination. This log-likelihood surface $\ell(N_e, t_{1/2})$ is given in its entirety in figure 5.6.

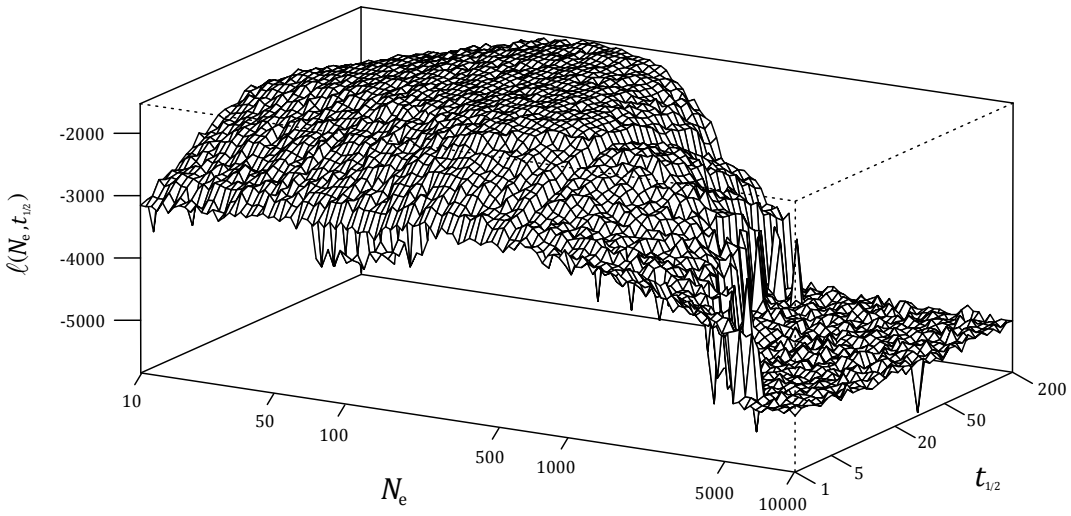
The peak of this surface, i.e. the MLE $(\hat{N}_e, \hat{t}_{1/2})$, lies at (100, 141), so the method exactly recaptures the known parameter values used to create our simulated dataset of 100 'observations'. We can construct a 95% confidence region for this estimate by taking just that part of the log-likelihood surface greater than:

$$\max\left\{\ell(N_e, t_{1/2})\right\} - \chi^2_{2,5\%}/2$$

which defines the region of parameter values that cannot be rejected by a likelihood-ratio test at the 5% level. So this 95% confidence region is given by the contour at:

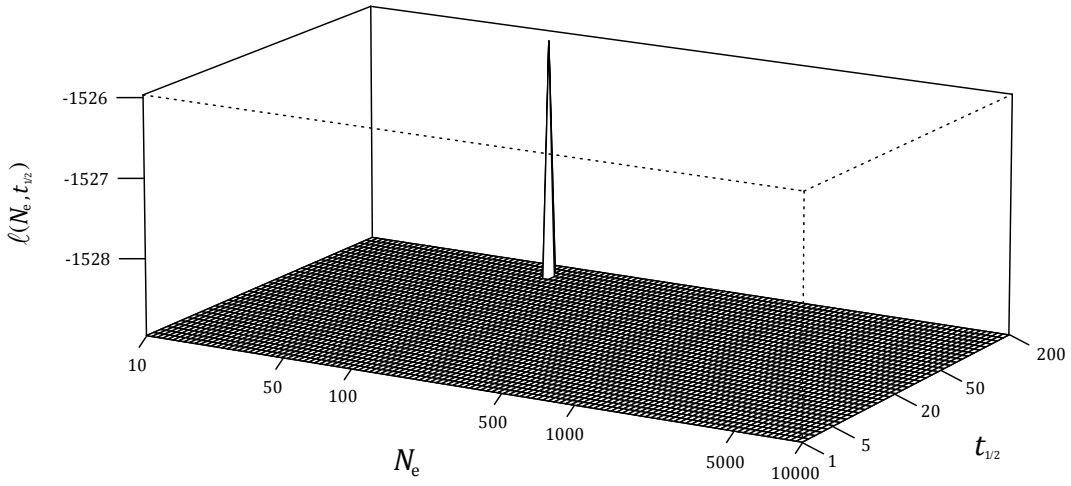
$$\ell(\hat{N}_e, \hat{t}_{1/2}) - 5.99/2 = -1525.966 - 2.996 = -1528.962$$

**Figure 5.6:** The full log-likelihood surface for $N_e = 100$, $t_{1/2} = 141$.




This region is shown in figure 5.7, and contains solely the point $(\hat{N}_e, \hat{t}_{1/2})$ demonstrating that the inference method appears accurate, i.e. every other parameter set can be rejected; although it is important to note that $N$ and $t_{1/2}$ values are on a grid, i.e. take discrete rather than continuous values.

**Figure 5.7:** The 95% confidence region for $N_e$ = 100, $t_{1/2}$ = 141.




**Test dataset 2: $N_e$ = 501, $t_{1/2}$ = 5**

Similarly, as a second example results for $N$ = 501, $t_{1/2}$ = 5 are given. Again the MLE $(\hat{N}_e, \hat{t}_{1/2})$ = (501, 5) recaptured the true parameter values from which the 'observed' data were simulated under. Similarly, the 95% confidence region is given by the contour at:

$$\ell(\hat{N}_e, \hat{t}_{1/2}) - 2.996 = \text{-}657.114 - 2.996 = \text{-}660.110$$

which is shown in figure 5.9, and again contains solely the point $(\hat{N}_e, \hat{t}_{1/2})$.

**Figure 5.8:** The full log-likelihood surface for $N_e = 501$, $t_{1/2} = 5$.

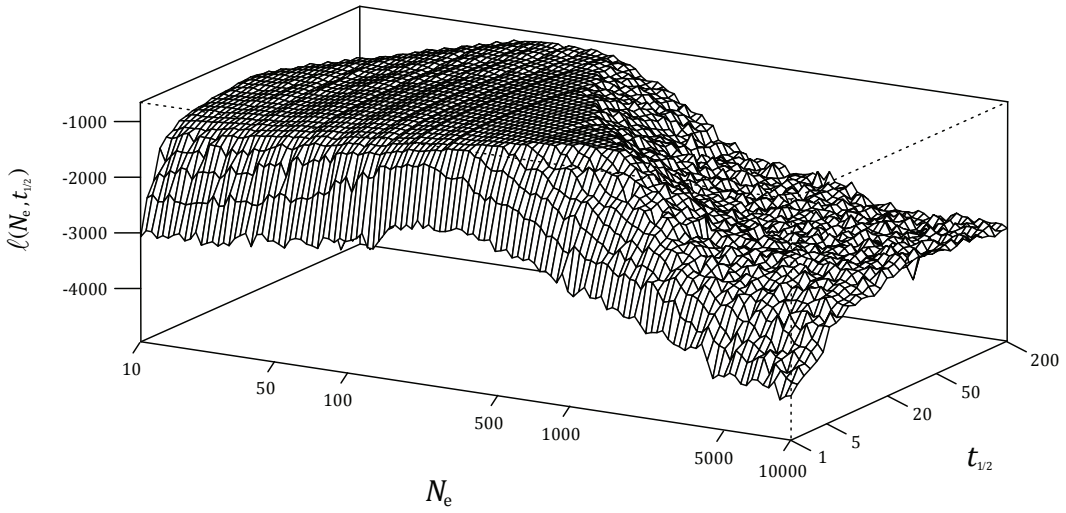
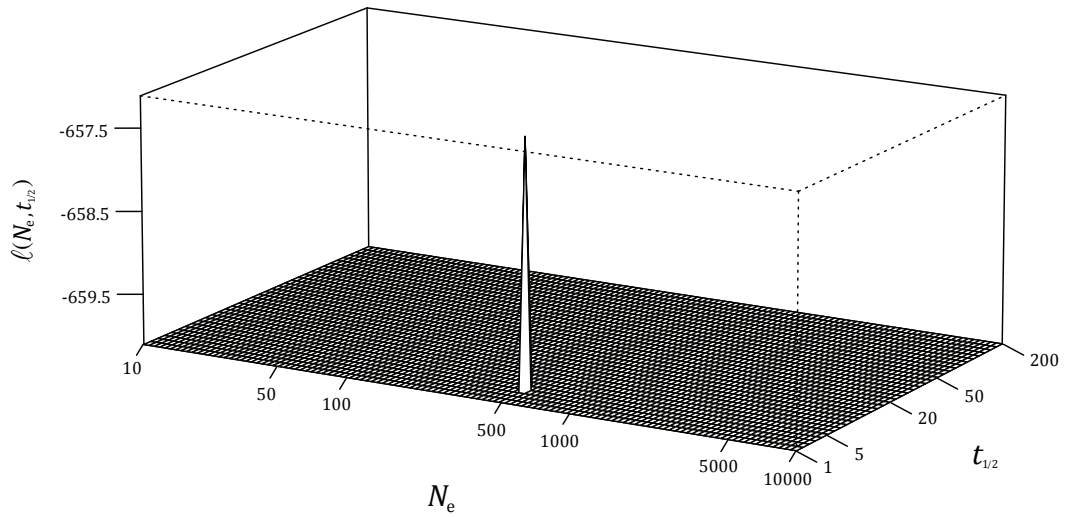



**Figure 5.9:** The 95% confidence region $N_e = 501$, $t_{1/2} = 5$.



Both results presented thus far were based on just **one** simulated 'observed' sample of size 100, and these single estimates of the true parameter values $(\hat{N}_e, \hat{t}_{1/2})$ were exact. The 95% confidence regions for both examples included solely these MLEs, so it is difficult to quantify uncertainty in the inference method in this way. One way to remedy this was to repeat the estimation of the MLE multiple times using a unique simulated 'observed' sample in every iteration. These multiple independent estimates $(\hat{N}_e, \hat{t}_{1/2})$ could then be used to create empirical 95% confidence regions for this joint estimate, as well as marginal densities, modal estimates and 95% confidence intervals (here denoted $CI_f$ to distinguish them from Bayesian credible intervals CI used elsewhere in this thesis) for each parameter $\hat{N}_e$ and $\hat{t}_{1/2}$ separately. The algorithm for this process is given in the next section.

### 5.4.4 Parameter estimation algorithm for test datasets

#### 5.4.4.1 The algorithm

Given that the full grid of estimated joint densities $f(\kappa, \tau)$ had been created, the full inference algorithm is as follows:

1. Independently simulate a large set of values $(\kappa_i, \tau_i)$ under the model with parameters $(N_e, t_{1/2})$.
2. Choose sample size *s* and number of iterations *it*.
3. Create *it* 'observed' datasets of size *s* by sampling the set of $(\kappa_i, \tau_i)$ values.
4. For each iteration calculate the MLE $(\hat{N}_e, \hat{t}_{1/2})$ as above.
5. Using all *it* iterations, empirically estimate a 95% confidence region for the joint MLE, and marginal modal estimates and 95% $CI_f$ for each parameter.

#### 5.4.4.2 Basic cultural drift model – results from the algorithm

Following application of this algorithm, the results for the five chosen test data sets are given below in table 5.2 and figures 5.10 – 5.14. The number of independent iterations *it* was set at 10,000 throughout, and sample size *s* was reduced to 50. Though test results with $s = 100$ were uniformly excellent, this reduced sample size of 50 was chosen as a conservative measure. Further, the archaeological dataset considered in the penultimate section is of this magnitude ($s = 42$), so it was useful to test how accurate and robust the method was with comparably sized simulated datasets.

**Table 5.2:** MLEs for effective population size $N_e$ and half-life $t_{1/2}$ for sample size $s = 50$

| | $\hat{N}_e$ | | $\hat{t}_{1/2}$ | |
|---|---|---|---|---|
| $(N_e, t_{1/2})$ | mode | 95% $CI_f$ | mode | 95% $CI_f$ |
| $N_e = 50$, $t_{1/2} = 28$ | 49.89 | [42.09, 56.38] | 26.46 | [20.49, 33.06] |
| $N_e = 100$, $t_{1/2} = 141$ | 100.77 | [89.45, 167.31] | 140.97 | [28.57, 179.61] |
| $N_e = 501$, $t_{1/2} = 5$ | 516.01 | [444.11, 642.35] | 5.06 | [3.64, 6.67] |
| $N_e = 1000$, $t_{1/2} = 70$ | 991.96 | [861.14, 1201.41] | 62.99 | [45.68, 87.13] |
| $N_e = 5011$, $t_{1/2} = 11$ | 4794.28 | [3627.35, 5517.76] | 10.12 | [7.44, 11.87] |

**Figures 5.10 to 5.14 (over the following pages):** Maximum likelihood estimates for $(N_e, t_{1/2})$. The 95% confidence region is given for the joint MLE as a contour, while densities for the marginal MLEs are given on each axis. The red point gives the true simulated parameter values.

**Figure 5.10:** MLEs for $N_e$ = 50, $t_{1/2}$ = 28.

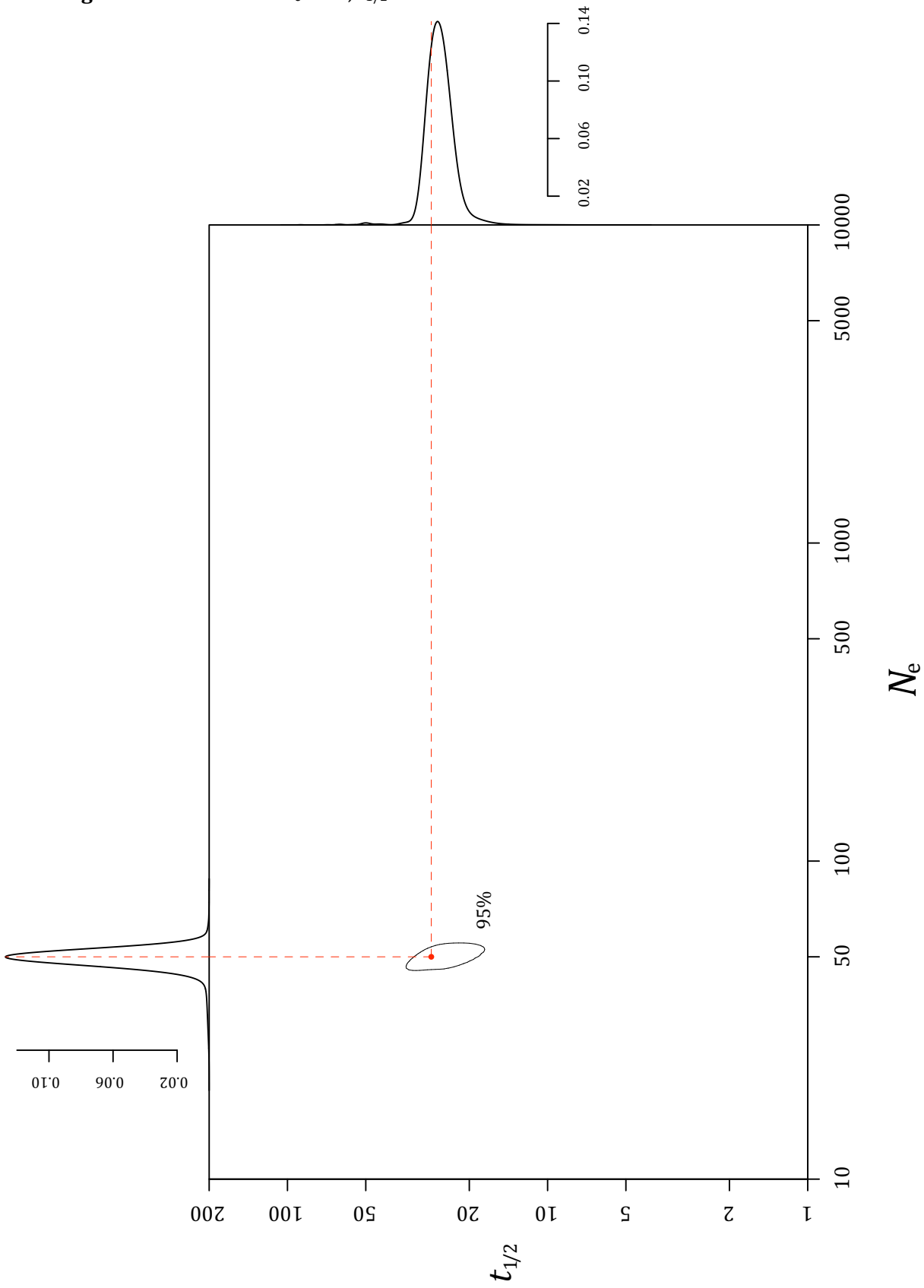


**Figure 5.11:** MLEs for $N_e$ = 100, $t_{1/2}$ = 141.

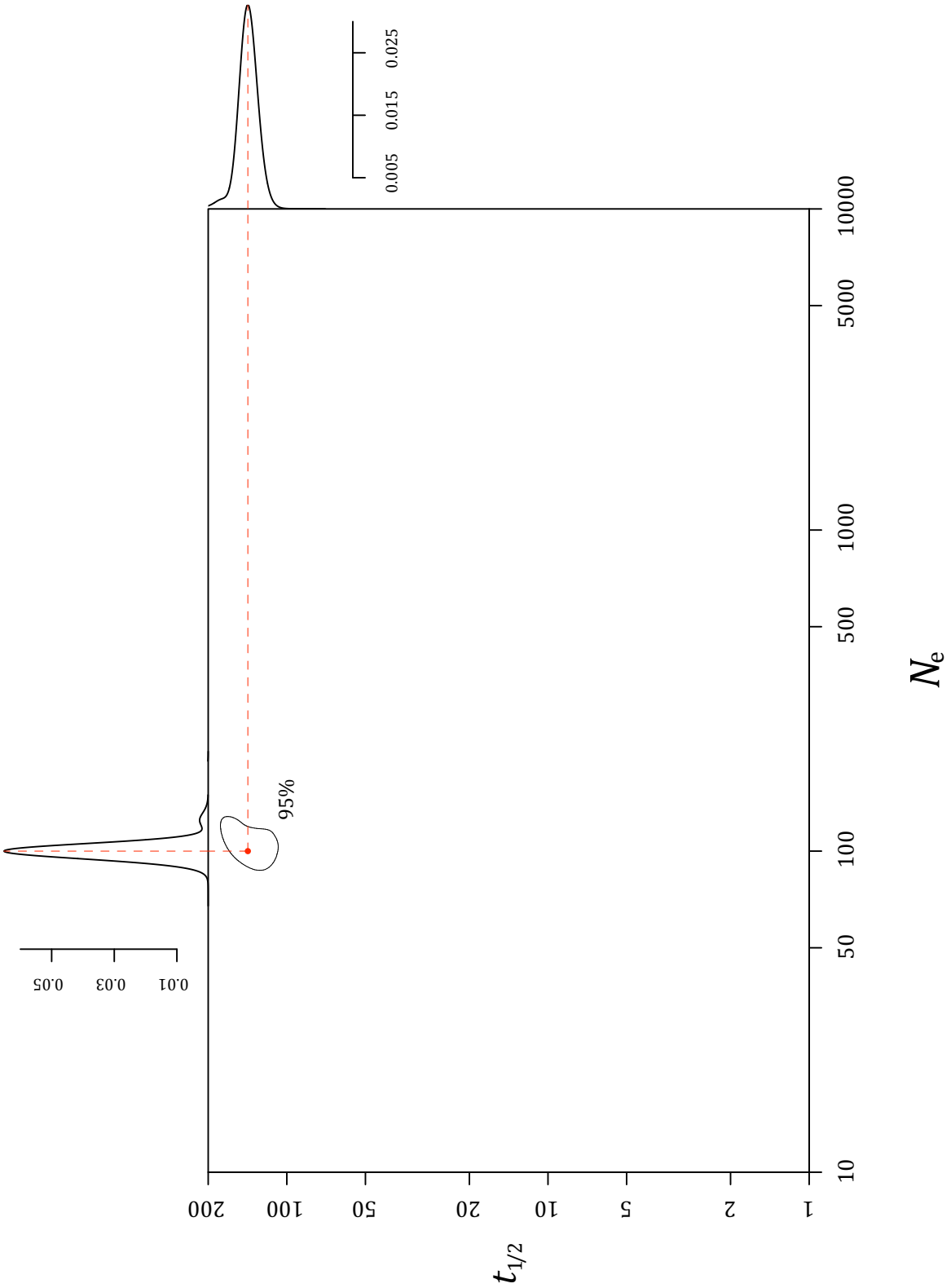


**Figure 5.12:** MLEs for $N_e = 501$, $t_{1/2} = 5$.

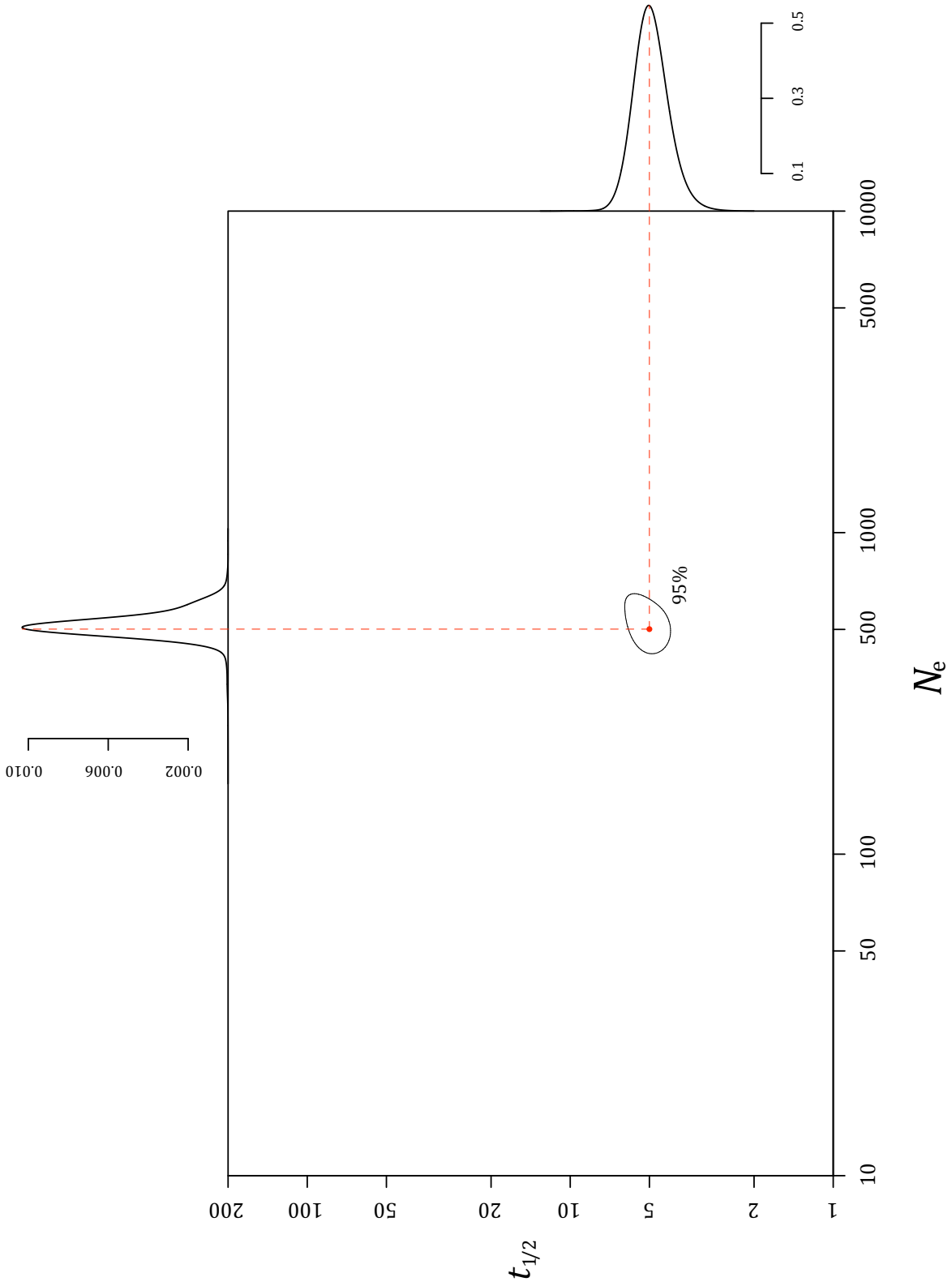


**Figure 5.13:** MLEs for $N_e$ = 1000, $t_{1/2}$ = 70.

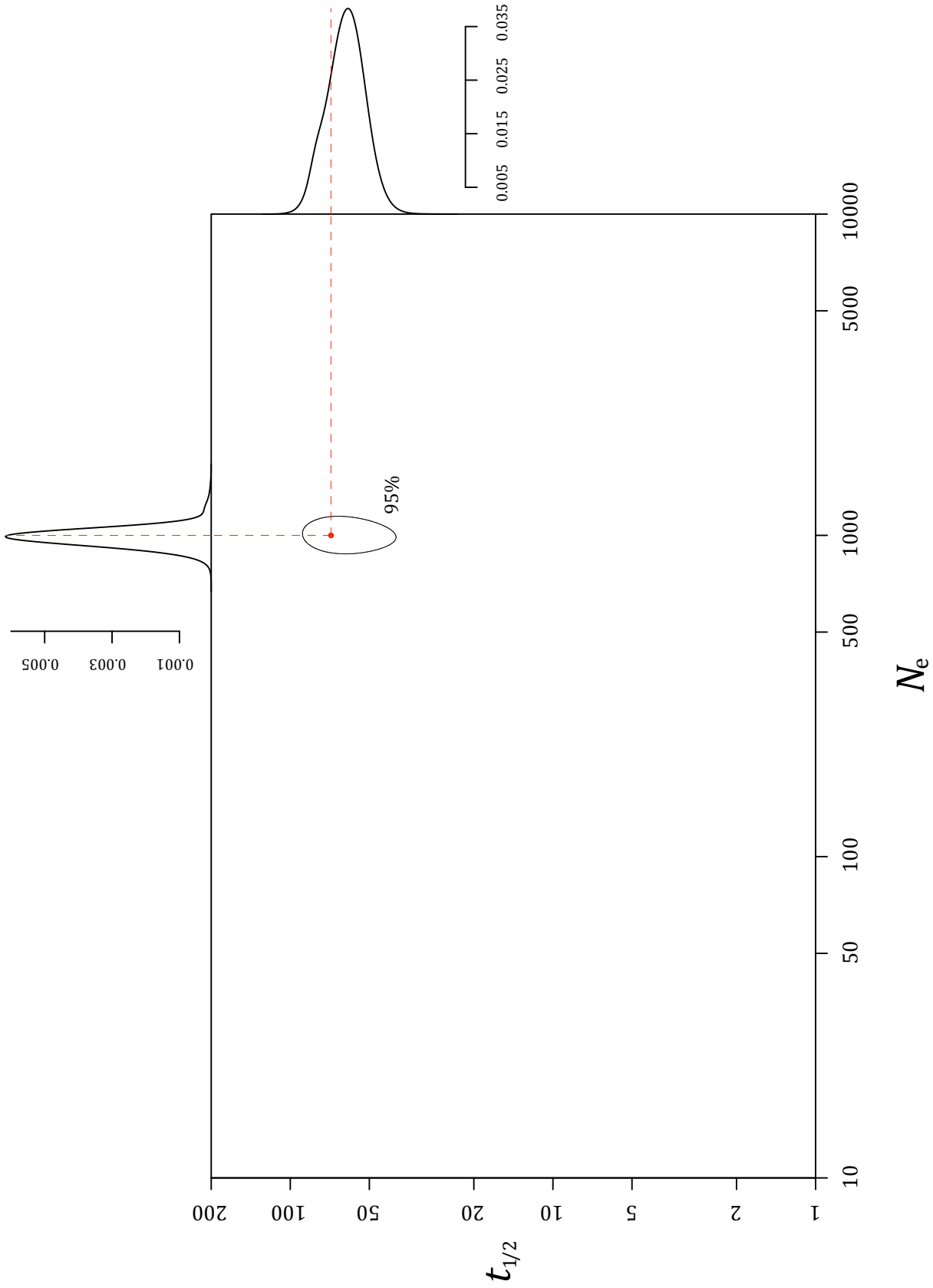
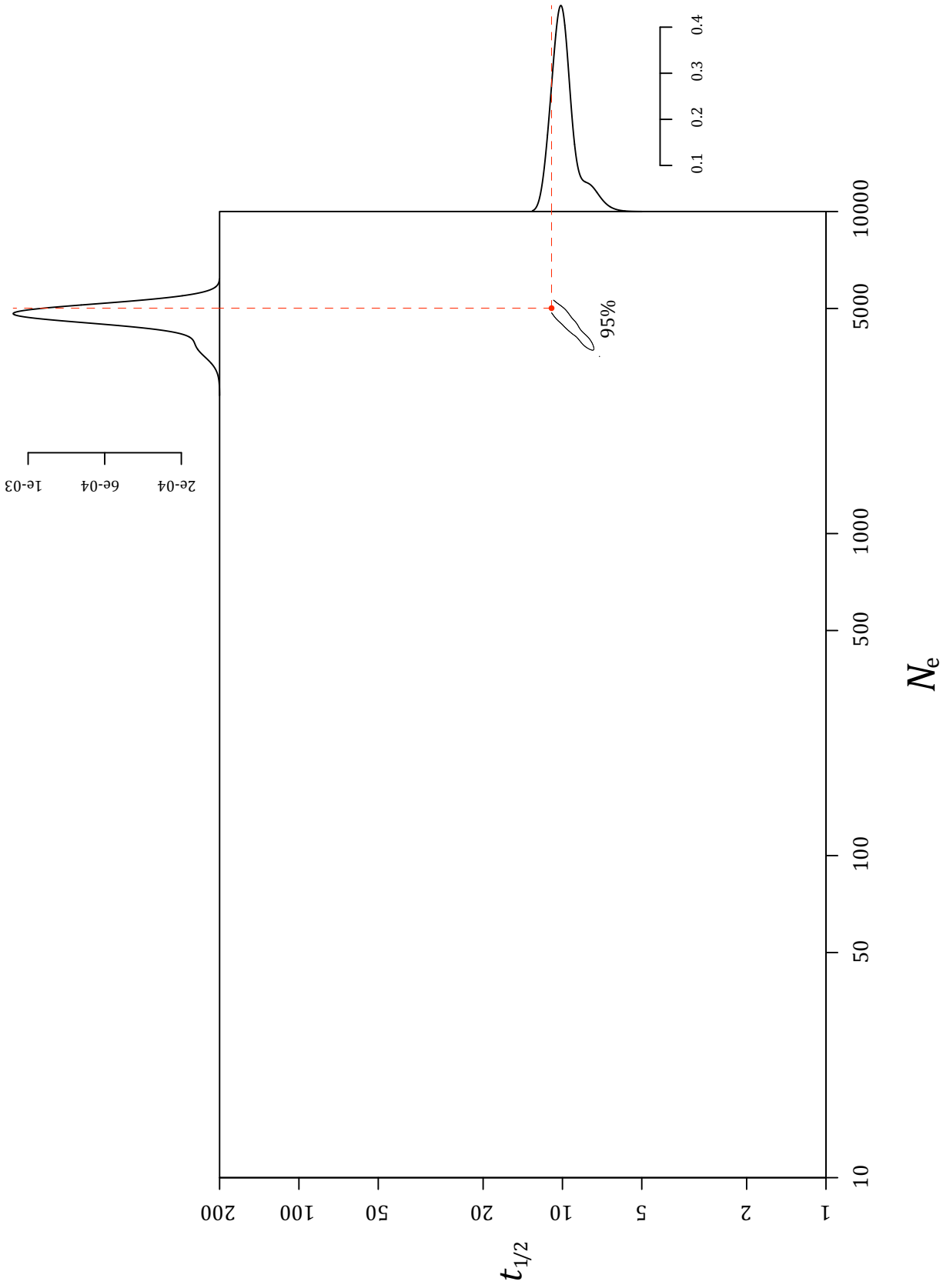


**Figure 5.14:** MLEs for $N_e$ = 5011, $t_{1/2}$ = 11.



All marginal parameter MLE modes appeared to be in excellent agreement with the true simulated parameter values, and all estimated 95% confidence intervals (see table 5.2) and 95% joint confidence regions (see figures 5.10 – 5.14) clearly contained the true simulated values. Whilst so far sample size $s$ had been set at 50, this inference method was able to accurately recapture the true parameters using far smaller (and more realistic) sample sizes. Indeed, the modal parameter estimates with $s$ = 10 were still very good (table 5.3), though the 95% $CI_f$s were broader, as would be expected. It should be noted that with $N_e$ = 100 and $t_{1/2}$ = 141, the 95% $CI_f$ for $\hat{t}_{1/2}$ was very wide. This may suggest difficulties in inference for this parameter when the half-life is extremely high and we only have a small amount of data available.

**Table 5.3**: MLEs for effective population size $N_e$ and half-life $t_{1/2}$ for sample size $s$ = 10

| | $\hat{N}_e$ | | $\hat{t}_{1/2}$ | |
|---|---|---|---|---|
| $(N_e, t_{1/2})$ | mode | 95% $CI_f$ | mode | 95% $CI_f$ |
| $N_e$ = 50, $t_{1/2}$ = 28 | 48.23 | [19.04, 74.49] | 28.50 | [10.10. 127.83] |
| $N_e$ = 100, $t_{1/2}$ = 141 | 121.38 | [89.86,183.73] | 138.55 | [7.79, 205.53] |
| $N_e$ = 501, $t_{1/2}$ = 5 | 602.39 | [283.85, 1149.56] | 4.70 | [1.99, 10.52] |
| $N_e$ = 1000, $t_{1/2}$ = 70 | 1190.91 | [841.34, 1909.98] | 54.13 | [29.68, 83.53] |
| $N_e$ = 5011, $t_{1/2}$ = 11 | 4781.19 | [3308.24, 5676.31] | 9.95 | [6.28, 12.01] |

In conclusion, this basic model appeared able to recover the true parameter values very well even with a very small sample of test 'observed' data ($s$ = 10). So far I had not incorporated the likely effects cultural innovation processes would have on variant frequencies, and in the next section this extension to the model is introduced.

### 5.4.5 The cultural drift model with innovation

#### 5.4.5.1 Cultural innovation

Following Neiman (1995, see also Chapter 4; section 4.4.1), I make the assumption that every innovation event results in a novel cultural variant. This is somewhat analogous to the infinite-alleles model of population genetics (see Kimura and Crow 1964), in which every mutation gives rise to a new allele. However, as this inference method relies on the joint probability density of the frequency trajectory of just **one** focal variant, any innovation that occurs anywhere else in the cultural pool is irrelevant for our purposes. The only concrete effect of innovation is to reduce the frequency of this focal variant, as focal variant artefacts are replaced with novel ones. Because I consider all elements of the pool of *non*-focal artefacts as interchangeable (and essentially uninteresting) I am really modeling just two mutually exclusive subsets: focal and non-focal, with innovation acting in one direction only.

As such the appropriate starting point is the two-allele Moran model with one-way mutation, i.e. allele $A_1$ can mutate into allele $A_2$ but not vice versa. Of course one consequence of using just one-way mutation is that the focal variant is guaranteed to go extinct eventually, but again this is of no concern here as I am not

interested in long-term equilibrium behaviour, but rather the maximum frequency reached and the time taken to reach this maximum. Again it is important to note that I am implicitly also making the assumption, as I have throughout this chapter, that all possible cultural variants are selectively neutral.

Given these assumptions and a rate of innovation per year $\mu_{\text{year}}$ estimated from the data, I first needed to scale this to a per-simulation time-step rate $\mu_{\text{step}}$ as follows:

$$\mu_{\text{step}} = \mu_{\text{year}} t_{\text{step}}$$

$$= \frac{\mu_{\text{year}} \ln\{N_{\text{e}} / (N_{\text{e}} - 1)\}}{\ln(2)} t_{1/2}$$

The parameter range remained fixed at $10 \leq N_{\text{e}} \leq 10{,}000$ and $1 \leq t_{1/2} \leq 199$, meaning that a simulation time-step $t_{\text{step}}$ could take values in the interval:

$$t_{\text{step}} = \left[ \min\left( \frac{\ln\{N_{\text{e}} / (N_{\text{e}} - 1)\}}{\ln(2)} t_{1/2} \right), \max\left( \frac{\ln\{N_{\text{e}} / (N_{\text{e}} - 1)\}}{\ln(2)} t_{1/2} \right) \right]$$

$$= \left[ \left( \frac{\ln\{10000/9999\}}{\ln(2)} \text{x } 1 \right), \left( \frac{\ln\{10/9\}}{\ln(2)} \text{x } 199 \right) \right]$$

$$= [0.0001,\ 30.25]$$

The number of innovations per time-step $\mu_{\text{step}}$ by definition has to be less than or equal to 1 (but ideally << 1) as only one artefact is replaced per time-step, which

constrained the absolute maximum value of innovation rate per year $\mu_{\text{year}}$ to:

$$\mu_{\text{step}} \leq 1$$

$$\mu_{\text{year}} \leq \frac{1}{\max(t_{\text{step}})} = 1/30.25...$$

$$\mu_{\text{year}} \leq 0.0331$$

This maximum constraint applied only to the *full* parameter set I have chosen for both $N_e$ and $t_{1/2}$ here. If I chose to restrict this parameter space then this maximum rate would increase. This means that if I increased the minimum effective population size $N_e$ and restricted $t_{1/2}$ to a smaller maximum, this model could comfortably deal with a higher yearly rate of innovation that might be more appropriate in some cultural contexts. In order to make clear that this is not a shortcoming of the method in general, imagine for example that we had a dataset for which a yearly innovation rate of 5% had been estimated. This would simply restrict the ranges of possible values for parameters $N_e$ and $t_{1/2}$ to the region where
$t_{\text{step}} = \{\ln(N_e/(N_e - 1))/\ln(2)\} t_{1/2} \leq 1/0.05 = 20$.

Indeed with low $N_e$ and high $t_{1/2}$ in general the value of $\mu_{\text{step}}$ can be very high – to the extent that innovation can no longer be considered a rare occurrence per simulation step. At this extreme, the amount of noise created by this high level of innovation will understandably disguise any signal of an underlying cultural evolutionary process, and indeed it can be questioned as to whether it constitutes an evolutionary system at all. This type of argument underlies many criticisms of cultural evolution in general (Atran 2001, Sperber 1996, Boyer 1999, but see also

Henrich and Boyd 2002 and Sperber and Claidiére 2008), but determining the rate of innovation in any system surely remains an empirical question, and the extent to which evidence for cultural evolutionary processes can be uncovered will depend critically on this.

For this reason, I restricted the *yearly* rates $\mu_{\text{year}}$, explored to 0.001, 0.005, 0.01, which depending on the values of $N$ and $t_{1/2}$ corresponded to rates per *time-step* $\mu_{\text{step}}$ in the ranges [1.4428 x10$^{-7}$, 0.0303], [7.2138 x10$^{-7}$, 0.1512] and [1.4428 x10$^{-6}$, 0.3025] respectively, i.e. scaled by the min and max $t_{\text{step}}$ values calculated above.

With the innovation rate $\mu_{\text{step}}$ so defined, the transition probabilities for this model are as follows where, again, $c$ is the count of the focal variant within the pool:

$$p_{c,c+1} = \frac{(1-\mu_{\text{step}})\,c\,(N_e - c)}{N_e\,(N_e - 1)}$$

$$p_{c,c-1} = \frac{\mu_{\text{step}}\left\{(N_e - c)^2 + c^2 - N_e\right\} + c\,(N_e - c)}{N_e\,(N_e - 1)}$$

$$p_{c,c} = \frac{(1-\mu_{\text{step}})\left\{(N_e - c)^2 + c^2 - N_e\right\} + \mu_{\text{step}}\left\{c\,(N_e - c)\right\}}{N_e\,(N_e - 1)}$$

Again 1 million simulations were carried out for each parameter combination $N_e$ and $t_{1/2}$, and for each of the three yearly innovation rates, $\mu_{\text{year}}$, 0.001, 0.005 and 0.01. Estimated joint densities $f(\kappa, \tau)$ were constructed using the process previously described, resulting in a grid of $f(\kappa, \tau)$ over $N_e$ and $t_{1/2}$ for each of these rates $\mu_{\text{year}}$.

#### 5.4.5.2 Cultural drift with innovation – results from the algorithm

Simulated 'observed' datasets were created for each of the five parameter combinations ($N_e$ , $t_{1/2}$) in a similar way as in the basic model above, except that I now simulated with the additional innovation parameter $\mu_{year}$, which was automatically scaled to the appropriate $\mu_{step}$. Results are given for sample size $s$ = 50 and 10 in tables 5.4 and 5.5 respectively, for each of the three values of $\mu_{year}$ 0.001, 0.005 and 0.01.

For sample size $s$ = 50, almost all modal estimates for all innovation rates again appear very good. The modal estimates for the $N_e$ = 100, $t_{1/2}$ = 141 test example are elevated, the possible reasons for which will be discussed below. Also at the higher innovation rates the modal estimate for $\hat{t}_{1/2}$ the $N_e$ = 1000, $t_{1/2}$ = 70 is slightly underinflated. The majority of the 95% $CI_f$s contain the true simulated values; only at the higher innovation rates do the $\hat{N}_e$ 95% $CI_f$ for $N_e$ = 100, $t_{1/2}$ = 141, and the $\hat{t}_{1/2}$ 95% $CI_f$ for $N_e$ = 1000, $t_{1/2}$ = 70, no longer contain the true simulated values.

**Table 5.4**: Sample size $s$ = 50. MLEs for effective population size $N_e$ and half-life $t_{1/2}$ with innovation rates per year $\mu_{year}$ = 0.001, 0.005 and 0.01.

| | | $\hat{N}_e$ | | $\hat{t}_{1/2}$ | |
|---|---|---|---|---|---|
| $(N_e, t_{1/2})$ | $\mu_{year}$ | mode | 95% $CI_f$ | mode | 95% $CI_f$ |
| $N_e = 50$ $t_{1/2} = 28$ | 0.001 | 49.89 | [38.67, 60.51] | 28.14 | [19.86, 52.57] |
| | 0.005 | 50.12 | [38.98, 68.75] | 27.87 | [17.81, 55.57] |
| | 0.01 | 50.38 | [39.66, 68.36] | 26.57 | [17.16, 42.16] |
| $N_e = 100$ $t_{1/2} = 141$ | 0.001 | 124.56 | [95.65, 138.05] | 174. 74 | [91.73, 208.63] |
| | 0.005 | 125.02 | [103.22, 137.09] | 178.50 | [94.93, 211.28] |
| | 0.01 | 124.93 | [109.89, 136.20] | 183.44 | [135.30, 215.26] |
| $N_e = 501$ $t_{1/2} = 5$ | 0.001 | 516.24 | [442.86, 641.22] | 5.07 | [3.63, 6.71] |
| | 0.005 | 515.72 | [442.79, 640.59] | 5.05 | [3.61, 6.68] |
| | 0.01 | 515.63 | [429. 29, 642.13] | 5.04 | [3.57, 6.75] |
| $N_e = 1000$ $t_{1/2} = 70$ | 0.001 | 994.37 | [863.34, 1248.62] | 64.00 | [44.99, 87.42] |
| | 0.005 | 991.96 | [857.54, 1365. 84] | 58.14 | [39.73, 80.54] |
| | 0.01 | 988.10 | [839.22, 1460.56] | 50.79 | [33.08, 68.74] |
| $N_e = 5011$ $t_{1/2} = 11$ | 0.001 | 4793.73 | [3644.85, 5539.09] | 10.10 | [7.42 11.83] |
| | 0.005 | 4762.16 | [3567.99, 5498.90] | 10.05 | [7.31, 11.81] |
| | 0.01 | 4665.78 | [3403.88, 5380.19] | 9.82 | [6.92, 11.58] |

The results are very similar for sample size $s$ = 10, except that, as expected, the $CI_f$s are in some cases considerably wider. This demonstrates that this estimation method is surprisingly robust to sample size, and that estimates are likely to be very reasonable with only very few observed archaeological samples.

**Table 5.5**: Sample size $s$ = 10. MLEs for effective population size $N_e$ and half-life $t_{1/2}$ with innovation rates per year $\mu_{year}$ = 0.001, 0.005 and 0.01.

| | | $\hat{N}_e$ | | $\hat{t}_{1/2}$ | |
|---|---|---|---|---|---|
| $(N_e, t_{1/2})$ | $\mu_{year}$ | mode | 95% $CI_f$ | mode | 95% $CI_f$ |
| $N_e$ = 50<br>$t_{1/2}$ = 28 | 0.001 | 47.56 | [20.12, 72.47] | 30.99 | [13.77, 140.73] |
| | 0.005 | 47.89 | [20.62, 72.79] | 30.71 | [13.13, 143.72] |
| | 0.01 | 48.84 | [21.20, 79.18] | 28.25 | [10.66, 140.99] |
| $N_e$ = 100<br>$t_{1/2}$ = 141 | 0.001 | 123.18 | [91.99, 140.36] | 174.28 | [72.20, 213.99] |
| | 0.005 | 124.25 | [95.44, 139.25] | 179.48 | [75.06, 215.29] |
| | 0.01 | 124.47 | [96.34, 139.23] | 183.63 | [79.35, 216.29] |
| $N_e$ = 501<br>$t_{1/2}$ = 5 | 0.001 | 596.85 | [282.52, 1114.56] | 4.74 | [2.03, 10.80] |
| | 0.005 | 605.32 | [294.20, 1132.17] | 4.62 | [1.99, 10.31] |
| | 0.01 | 602.22 | [271.20, 1179.42] | 4.62 | [1.91, 11.32] |
| $N_e$ = 1000<br>$t_{1/2}$ = 70 | 0.001 | 1234.52 | [852.42, 1929.61] | 52.96 | [28.32, 82.98] |
| | 0.005 | 1327.50 | [865.83, 2011.77] | 45.77 | [21.51, 74.89] |
| | 0.01 | 1393.31 | [845.36, 2069.59] | 37.62 | [14.87, 66.22] |
| $N_e$ = 5011<br>$t_{1/2}$ = 11 | 0.001 | 4796.39 | [3301.20, 5713.76] | 9.97 | [6.20, 12.01] |
| | 0.005 | 4752.31 | [3299.00, 5655.93] | 9.87 | [6.23, 11.98] |
| | 0.01 | 4631.83 | [3231.36, 5593.04] | 9.52 | [5.98, 11.87] |

There are two possible reasons for this seemingly anomalously inflated $N_e$ = 100, $t_{1/2}$ = 141 result, the first of which was hinted at above (section 5.4.4.2). This is the fact that with a low population size and high half-life, the duration of each simulation time-step (i.e. the scaling factor $t_{step}$) will be large, so it follows that the per step innovation rate $\mu_{step}$ will be high. For this example with $N_e$ = 100 and $t_{1/2}$ = 141, the value of $t_{step} = \{\ln(100/99)/\ln(2)\}.141 = 2.044$ years. So innovation rates per year $\mu_{year}$ of 0.005 and 0.01 would be converted to innovation rates per step $\mu_{step}$ of 0.0122

and 0.02044 respectively. Table 5.6 gives the three per year innovation rates $\mu_{year}$ scaled to the appropriate $\mu_{step}$ values for each of the 5 test datasets. Compared to the $\mu_{step}$ values for the other 4 test examples these values are very high (* cells in table 5.6), and this may constitute part of the problem.

**Table 5.6:** Innovation rates per simulation step $\mu_{step}$ for innovation rates per year $\mu_{year}$ for the five simulated test datasets.

| | | $\mu_{step}$ | | |
|---|---|---|---|---|
| $(N_e, t_{1/2})$ | $t_{step}$ | $\mu_{year} = 0.001$ | $\mu_{year} = 0.005$ | $\mu_{year} = 0.01$ |
| $N_e = 50, t_{1/2} = 28$ | 0.8160 | 0.00082 | 0.00408 | 0.00816 |
| $N_e = 100, t_{1/2} = 141$ | 2.0444 | 0.00204* | 0.01022* | 0.02044* |
| $N_e = 501, t_{1/2} = 5$ | 0.0144 | 0.00001 | 0.00007 | 0.00014 |
| $N_e = 1000, t_{1/2} = 70$ | 0.1010 | 0.00010 | 0.00051 | 0.00101 |
| $N_e = 5011, t_{1/2} = 11$ | 0.0032 | 0.00000 | 0.00002 | 0.00003 |

The second more prosaic, yet also more likely reason may be that with such a high half-life (and thus a high $t_{step}$ value) the simulation model was not capturing sufficient time-steps of data; recall that the time-axis on each estimated joint density $f(\kappa, \tau)$ was truncated to 500 years (for computational efficiency). This means that with $t_{step} = 2.0444$ only maximum variant counts ($c_{max}$) which occurred during the first $\sim$250 time-steps of each of the original 1 million simulations would be used to construct the joint densities. This can very probably be remedied by either extending the time range over which each of the joint densities $f(\kappa, \tau)$ is constructed, or by

limiting the half-life to a sensible range. The possible range of cultural half-lives, as defined here, and the duration over which cultural evolutionary processes themselves may endure, is entirely context dependent, and this method will need to be carefully calibrated in order to get the best out of any specific archaeological dataset.

This discrepancy aside, however, the method appears to work very well. The above results amply demonstrate that if we are able to estimate the yearly innovation rate from our observed data and use this estimate to condition the inference algorithm in the way described above, then the parameter estimates are very accurate.

#### 5.4.5.3 Sensitivity to innovation

However, a further observation, drawn from table 5.6, was that for a given $\mu_{\text{year}}$ rate the scaled $\mu_{\text{step}}$ values can be very small indeed – especially for large $N_{\text{e}}$ and small $t_{1/2}$ – so it seemed sensible to test the sensitivity of this method to innovation. What if we were unable to provide an empirical innovation rate estimate? What rate of innovation in the data-generating process would impact our ability to make accurate inferences, i.e. what is the upper limit for innovation which we can 'ignore'?

In order to start to answer these questions I generated test samples as before for the three previously chosen yearly innovation rates $\mu_{\text{year}}$, 0.001, 0.005 and 0.01, and then used them to estimate the MLEs $(\hat{N}_{\text{e}}, \hat{t}_{1/2})$ as previously described, *but* using the original grid of joint densities $f(\kappa, \tau)$ from the basic model with *no innovation*. I provide results below for four of the previous test parameter sets with sample size, *s*

= 50, omitting the $N_e$ = 100, $t_{1/2}$ = 141 test for the reasons discussed above. All results presented in this section used 5,000 independent iterations of MLE estimation to derive the modal values and 95% $CI_f$s.

**Table 5.7**: Sample size $s$ = 50. MLEs for innovation rates per year $\mu_{year}$ = 0.001, 0.005 and 0.01, with likelihoods calculated from joint densities $f(\kappa, \tau)$ with no innovation.

| | | $\hat{N}_e$ | | $\hat{t}_{1/2}$ | |
|---|---|---|---|---|---|
| $(N_e, t_{1/2})$ | $\mu_{year}$ | mode | 95% $CI_f$ | mode | 95% $CI_f$ |
| $N_e$ = 50 $t_{1/2}$ = 28 | 0.001 | 49.86 | [38.55, 58.96] | 28.21 | [20.92, 54.77] |
| | 0.005 | 50.12 | [39.10, 69.18] | 27.17 | [16.33, 39.62] |
| | 0.01 | 50.70 | [39.18, 71.38] | 25.71 | [14.29, 37.21] |
| $N_e$ = 501 $t_{1/2}$ = 5 | 0.001 | 516.09 | [444.68, 639.58] | 5.04 | [3.62, 6.62] |
| | 0.005 | 515.82 | [444.04, 648.48] | 5.05 | [3.61, 6.68] |
| | 0.01 | 519.25 | [447.38, 652.83] | 5.03 | [3.54, 6.61] |
| $N_e$ = 1000 $t_{1/2}$ = 70 | 0.001 | 991.76 | [858.94, 1249.62] | 61. 18 | [43.37, 84.97] |
| | 0.005 | 979.36 | [830.33, 1351.64] | 55.71 | [37.41, 77.65] |
| | 0.01 | 974.29 | [807.60, 1430.23] | 46.09 | [29.13, 66.15] |
| $N_e$ = 5011 $t_{1/2}$ = 11 | 0.001 | 4794.28 | [3630.85, 5528.93] | 10.09 | [7.42, 11.83] |
| | 0.005 | 4741.21 | [3496.76, 5480.05] | 9.99 | [7.16, 11.79] |
| | 0.01 | 4730.73 | [3477.91, 5484.94] | 9.94 | [7.07, 11.73] |

Again all modal estimates in table 5.7 appeared very good; indeed, they were near indistinguishable from those produced from the full innovation model (table 5.4). The results for $s$ = 10 are given in table 5.8 for the highest innovation rate $\mu_{year}$ =

0.01 only, as the results for the two lower rates (0.001 and 0.005) were effectively identical to the full innovation model results given previously.

**Table 5.8**: Sample size $s$ = 10. MLEs for innovation rate per year $\mu_{\text{year}}$ = 0.01, with all likelihoods calculated from joint densities $f(\kappa, \tau)$ with no innovation.

| | | $\hat{N}_e$ | | $\hat{t}_{1/2}$ | |
|---|---|---|---|---|---|
| $(N_e, t_{1/2})$ | $\mu_{\text{year}}$ | mode | 95% $CI_f$ | mode | 95% $CI_f$ |
| $N_e = 50$ $t_{1/2} = 28$ | 0.01 | 48.98 | [21.53, 73.26] | 27.93 | [11.00, 126.28] |
| $N_e = 501$ $t_{1/2} = 5$ | 0.01 | 624.34 | [322.94, 1187.19] | 4.52 | [1.83, 9.51] |
| $N_e = 1000$ $t_{1/2} = 70$ | 0.01 | 1369.01 | [840.77, 2076.10] | 36.67 | [14.58, 62.38] |
| $N_e = 5011$ $t_{1/2} = 11$ | 0.01 | 4720.40 | [3246.71, 5676.87] | 9.70 | [5.98, 11.96] |

Again, understandably, with this smaller sample size the 95% $CI_f$s were wider but most modal estimates were still mostly excellent. However, the modes for the $N_e$ = 1000, $t_{1/2}$ = 70 test were overinflated for $\hat{N}_e$ and underinflated for $\hat{t}_{1/2}$. This is likely due to the relatively high-half life coupled with the lack of information in a sample of this small size. So with a sample of size $s$ of 50, and even as small as 10, I was still able to recapture the true parameter values with surprising accuracy even if I *ignored* the innovation occurring (at rate $\mu_{\text{year}} \leq 1\%$) in the process generating the sample.

To answer the second part of the question, i.e. 'what was the upper limit for innovation which we can 'ignore'?, I generated samples with rates of innovation $\mu_{\text{year}}$ incrementally increasing up to 10% for each of the 4 previous test parameter sets. Again, I then used the original grid of joint densities $f(\kappa, \tau)$ from the basic model with

*no innovation* to estimate the MLEs $(\hat{N}_e, \hat{t}_{1/2})$ for each innovation rate. For each parameter set I then found the maximum innovation rate $\mu_{year}$ for which the MLE modes were still reasonable, which gave a measure of the level of insensitivity of this method to the innovation process.

For sample size $s$ = 50, table 5.9 compares the MLE modes for each parameter estimated with the basic model from: (1) simulated test data with $\mu_{year}$ = 0 (from table 5.2); and (2) simulated test data with the maximum $\mu_{year}$ > 0 that still gave reasonable parameter estimates. As can be seen, for all parameter sets almost all the MLE modes generated from sample data *with* innovation are very close to the MLE modes from the original no-innovation model. The single discrepancy in the results was in the estimate of $\hat{t}_{1/2}$ for the parameter set $N_e$ = 1000, $t_{1/2}$ = 70, which may be for the reasons previously mentioned.

**Table 5.9**: The upper limit of innovation rate $\mu_{year}$ that can be 'ignored' in parameter estimation. MLEs for $\mu_{year}$ = 0 are given for comparison. Sample size $s$ = 50.

| $(N_e, t_{1/2})$ | $\mu_{year}$ | $\hat{N}_e$ mode | $\hat{t}_{1/2}$ mode |
|---|---|---|---|
| $N_e$ = 50, $t_{1/2}$ = 28 | 0 | 49.89 | 26.46 |
| | 0.02 | 53.55 | 19.26 |
| $N_e$ = 501, $t_{1/2}$ = 5 | 0 | 516.01 | 5.06 |
| | 0.05 | 524.10 | 4.86 |
| $N_e$ = 1000, $t_{1/2}$ = 70 | 0 | 991.96 | 62.99 |
| | 0.02 | 976.84 | 34.38 |
| $N_e$ = 5011, $t_{1/2}$ = 11 | 0 | 4794.28 | 10.12 |
| | 0.02 | 4662.29 | 9.78 |

These results, while not exhaustive, demonstrate that when the innovation rate estimated from the observed data is ≤2%, very reasonable MLEs for the parameters $N_e$ and $t_{1/2}$ can be obtained by using the simple no-innovation model. The results for $N_e = 501$, $t_{1/2} = 5$ suggest that for some parameter combinations the basic estimation method may be insensitive to up to 5% innovation per year. However, for real datasets and in the absence of any information on the parameters it is conservative to assume that when the estimated innovation rate is >2% per year that a full innovation model should be used.

### 5.4.6 Cultural drift model conclusions

In conclusion, this neutral model demonstrates that it is possible to model a cultural evolutionary process, without relying on the idea of 'cultural generations' and relaxing the assumption of 'cultural equilibrium'. Coupled with the estimation algorithm, this model provides very good approximate maximum likelihood estimates of the parameters $N_e$ and $t_{1/2}$ from simulated datasets. These estimates can be made with a realistically sized sample of 50, and remain accurate in most cases with sample sizes as small as 10. While a full innovation model has been developed, the last section demonstrated that it is not necessary to consider innovation if the rate estimated from the data is less than 2% per year. However, when this estimated rate is greater than 2% it appears advisable to use this rate paired with a grid of joint densities $f(\kappa, \tau)$ constructed appropriately with this yearly innovation rate for the calculation of approximate likelihoods. In the next section, I outline a previous neutral

model analysis of a dataset (Shennan and Bentley 2007, see also Chapter 4; section 4.4.2), before giving the results obtained with the model developed here.

## 5.5 Application to LBK pottery data

### 5.5.1 Overview of LBK data

The Linearbandkeramik (LBK) culture is associated with some of the earliest farming groups of central Europe (see Chapter 2; section 2.7.3), originating in modern-day northwest Hungary and southwest Slovakia around 7,500 ya before spreading rapidly westwards across Europe. The LBK is one of the most intensively studied pre-historical European cultures, and some evolutionary archaeological analyses of LBK data utilizing the neutral model have been outlined in the previous chapter (Chapter 4; sections 4.4.2 and 4.4.5, see also example Shennan and Wilkinson 2001, Bentley and Shennan 2004, Bentley *et al.* 2004).

The work of Strien (2000) provided a set of regional data on decorated pottery from the LBK settlements of southwest Germany. This dataset comprises of vessels from 270 separate excavation pits across the region bearing a total of 42 different decorative motifs, and was divided into 9 temporal phases. Strien's seriation analysis demonstrated a common chronological trend in the change of decorative motifs across the region, which indicated that the region may be considered as a single, interacting cultural population.

### 5.5.2 Previous analysis results

Shennan and Bentley (2007) previously analyzed this data utilizing the neutral drift model, a method described at length in Chapter 4. While Strien (2000) identified 9 phases, Shennan and Bentley aggregated phases 2B and 3 – as each phase had very small sample size – to give 8 phases in total. Using Strien's result that the whole region could be considered as a single population, Shennan and Bentley combined the motif frequency counts in each phase and from each pit to produce an aggregated frequency for each of the 42 motif variants in each phase for the whole southwest German region.

They then showed that the diversity estimators $t_E$ and $t_F$ (see Chapter 4; sections 4.4.1 and 4.4.2) for the whole region were consistent with the assumption of a neutral process / cultural drift, and that, as expected, both diversity estimators increased as population size increased (as measured by phase assemblage size standardized by phase length). They also demonstrated that the distribution of regional variant frequencies in each of the 8 phases were well described by a power-law, which they interpreted as further evidence of a neutral process (but see Chapter 4; section 4.5). Shennan and Bentley also made a direct estimation of the innovation rate by counting the number of novel variants that appeared in each phase and standardizing this count by the number of vessels in the phase and the phase length. They showed that the innovation rate appears to increase over the first 5 phases before crashing in the 6$^{th}$ phase, after which innovation almost completely ceases for the remainder of the duration of the LBK culture in the region.

### 5.5.3 Cultural drift model parameter estimation

I converted the 8 phase durations used by Shennan and Bentley (table A1 in the appendix) into years (1 'house generation' ≈ 25 years), to estimate the times from first appearance to maximum frequency for each of the 42 decorative variants. The maximum variant frequencies and these estimated times to maximum frequency ($\kappa_i$ , $\tau_i$) are given for each variant in table A2 in the appendix.

While Shennan and Bentley (2007) showed that innovation rates varied across the sites and across phases, all estimated rates were <2% *per generation*. When converted to years, all innovation rate estimates were clearly lower than the 2% *per year* identified above (section 5.4.5.3) as the critical value below which the parameters can be reasonably estimated without having to take innovation into account. This provides justification for using the basic no-innovation model to estimate parameter values $N_e$ and $t_{1/2}$.

For this dataset the approximate MLEs are $\hat{N}_e = 37$, $\hat{t}_{1/2} = 79$. These results suggest that the effective size of the pool of decorated vessels available to copy was very small, and not subject to a high rate of turnover – with a vessel only being replaced every $t_{step} = \{\ln(37/36)/\ln(2)\}.79 = 3.123$ years. The relationship between effective population and census population size is not always straightforward (see Chapter 1), but in general the effective size is substantially smaller. Under the Moran model this inferred effective population size would imply a census size $N$ of $2N_e = 74$ (see Felsenstein 1971a).

This is, of course, an idealized model and these results rely on the assumption of neutral cultural drift with regard to the propagation of vessel motifs. An alternative explanation for the low level of observed diversity could be that conformist copying-biases were restricting the spread of novel innovations throughout the region. However, the results of Shennan and Bentley (2007; figure 2) demonstrated that the regional variant frequencies are approximately power-law distributed, which they interpreted as evidence for a neutral random-copying process. My work on power-law distributions in the previous chapter (Chapter 4; section 4.6) showed that the introduction of copying-biases into an otherwise neutral process can lead to variant frequency distributions statistically indistinguishable from those generated through complete neutrality. This allows for the possibility that conformist copying was occurring, but in principle an upper-bound for the level of such conformity could be identified (see Chapter 4; section 4.6.3).

It should again be noted that this method currently makes the assumption that the cultural pool is of fixed size for the entire duration of artefact production. However, there is no reason in principle why this approach may not in future work be extended in order to relax this assumption.

## 5.6 Conclusion

### 5.6.1 Model conclusions

The model demonstrates that the assumption of cultural equilibrium is not required in order to accurately estimate the effective population size $N_e$ and half-life

$t_{1/2}$ of cultural artefacts from simulated data. Further, only a very minimal observed dataset is required for this estimation procedure – the maximum frequency and the time to maximum frequency for a small number of different cultural variants (possibly as few as 10) within a single assemblage. As has been mentioned, these are typically very easily estimated from archaeological data, which gives great scope for this model / estimation process.

By developing a full model that incorporates innovation, the above results have demonstrated that it is not *necessary* to take innovation into account if the rate estimated from the observed archaeological data is ≤2% per year. However, if the estimated rate is larger than this critical value, it is advisable to use the full innovation model, incorporating this estimated rate into calculation of likelihoods.

### 5.6.2 Testing cultural neutrality

This method very explicitly deals with neutral variants and their frequency trajectory under drift with various levels of cultural innovation. While any extension to deal with cultural selective forces in an explicit way would require wholesale changes – including the use or development of statistics that capture the full variant frequency spectrum rather than just the one, essentially random, focal variant – this method could also be used as a test for neutrality (see Chapter 4). With empirical estimates for $N_e$ and $t_{1/2}$ of the cultural pool from which a dataset derives, this method could be easily adapted to test for any deviation from that expected under neutrality, and thus allow either acceptance or rejection of simple cultural drift.

# Chapter 6

# Inferring demographic history: integrating ancient DNA

## 6.1 Summary

Recent advances in population genetics and associated statistical methods have allowed the integration of ancient DNA (aDNA) into a robust framework for inferring the demographic history of a population. The inclusion of aDNA in these coalescent-based analyses has been shown to improve inference on the timing and magnitude of past demographic events. The increasing availability of reliable aDNA sequence data, from a growing number of different species, has rapidly made this approach invaluable to population geneticists and archaeologists alike.

In this chapter I provide the background to coalescent theory and its application to inferring past demography using both modern and ancient DNA. I then give details of my work, as part of two separate collaborative projects. An analysis of ancient and modern Near Eastern cattle data provides a further line of evidence on the process of domestication, augmenting the evidence from archaeology. The analysis of Hispaniolan hutia genetic data may help to uncover its complex population (and taxonomic) history, which should prove useful for conservation purposes.

## 6.2 Abstract

In this chapter I first give an informal derivation of the basic coalescent approximation to the Wright-Fisher model (section 6.4.1). I then provide an outline of the many extensions that have subsequently been made to the coalescent model (section 6.4.2), before explaining how these provide an efficient method for generating simulated genetic data under a demographic model (section 6.4.3). I give a brief description of some of the current available software (6.4.5), and provide details of the application of approximate Bayesian computation (ABC) (described in depth in Chapter 2; section 2.5) to compare competing demographic models.

The final section (6.5) provides details of my application of this approach to inferring demographic history as part of collaborative projects on the Hispaniolan hutia (section 6.5.1) and Near Eastern taurine cattle (section 6.5.2). These analyses include using ABC for demographic model comparison, and employing the method of Voight *et al.* (2005) to provide validation for model choice.

## 6.3 Introduction

### 6.3.1 An overview of the coalescent model

Population genetic models have classically been concerned with predicting how allele frequencies in a population will change over time (Fisher 1930, Wright 1931, Moran 1958, 1962) by making various simplifying assumptions about how the genetic processes operate within that idealized population. These *prospective* models (Ewens 2004) based on 'population thinking' (Mayr 1982) make predictions – often derived through diffusion approximations – about the effects of population size or substructure, migration, or natural selection on the genetic variation within a population. The basics of two such models, those of Wright-Fisher and Moran, have been briefly illustrated in previous chapters (Chapter 4; section 4.4.1 and Chapter 5; section 5.4.2 respectively).

However, an alternative way of modelling genetic evolution within a population is based on 'genealogical thinking', and concerns itself with *retrospectively* determining the ancestral relationships between a contemporary sample of individuals. This type of thinking dates back at least to the work of Malécot during the 1940's (see for example Malécot 1948) on the 'identity by descent' of two sampled individuals, which was further generalized to more than just a pair of samples (see Nagylaki 1989 and also Griffiths 1980). One major advantage of this approach is that the observed sample is prioritized and the ancestry of non-sampled individuals can be disregarded. As will be seen below, centering the model on just the sampled data rather than a whole population can vastly increase the efficiency of analyses.

Kingman's work in the early 1980's (1982a, 1982b, though also see Hudson 1983, Tajima 1983, and Tavaré 1984) provided the first explicit link between the two existing lines of thought – 'population' and 'genealogical' – though numerous eminent predecessors laid important and necessary groundwork (see for example Felsenstein 1971b, Ewens 1972, Watterson 1975, Griffiths 1980). Kingman demonstrated that, given some important conditions, the genealogy of a sample of $n$ individuals taken from a population of size $N$ can be approximated by a continuous-time ancestral Markov process, which he termed the $n$-coalescent. In less abstract language, the $n$-coalescent consists of a rooted bifurcating tree structure of the ancestral relationships of the $n$ samples and the set of times for each of the $n - 1$ internal nodes of this genealogy.

The major assumptions are that any genetic variation is selectively neutral and that the population is panmictic, i.e. there is no population substructure. In effect this means that the number of offspring all individuals can have is identically distributed. But, while identical they are not independent as a further assumption is that the population size $N$ is fixed through time, so the total number of offspring has to sum to $N$. Joint distributions of this kind are called 'exchangeable' (see Cannings 1974), meaning that individuals or lineages in the ancestral tree can be labelled at random. The assumption of selective neutrality is important as it means that the genealogical process can be modelled as a purely population-based process entirely independently of the genetics of the constituent individuals, which is one of the greatest benefits of this approach (see section 6.4.3). Formally the coalescent approximation holds in the limit as $N$ tends to infinity with time scaled to units of $N$.

In practical terms this approximation remains reasonable when the sample size is small compared to a large population size, i.e. $n << N$.

In order to keep the following outline (section 6.4) of the coalescent model brief (I drop the *n*- for convenience) I provide it in terms of the haploid Wright-Fisher model, which was outlined in Chapter 4; section 4.4.1. It should be noted that the approximation holds for the diploid case (in the absence of recombination) and, given the appropriate time-scaling, for a broad category of genetic models (Cannings 1974, see also Wakeley 2009; pages 59-74) of which the Wright-Fisher is simply a special case. Rather than an approximation, the coalescent is an exact result for the Moran model, which was outlined in Chapter 5; section 5.4.2, but I do not give details here (see Wakeley 2009 for more on this).

### 6.3.2 The coalescent and ancient DNA

The development of the coalescent model (Kingman 1982a, 1982b) allows inferences to be made about the demographic history of a population from a relatively small number of genetic samples. A recent advance (Rodrigo and Felsenstein 1999) further allows the integration of ancient DNA (aDNA) samples into these model-based analyses. This is DNA that has been recovered from archaeological (or historical) specimens. The age of these specimens poses a number of methodological problems for the extraction and amplification of high quality sequence data, not least degradation and modern contamination, but see Hofreiter *et al.* (2001), Pääbo *et al*. (2004), Willerslev and Cooper (2005), and Gilbert *et al.* (2005) for technical reviews of how these can be minimized. The inclusion of aDNA can

greatly increase the power of demographic modelling approaches (see Drummond *et al*. 2003), and has been used to make inferences about the population history of a range of species (Shapiro *et al.* 2002, Chan *et al.* 2006, review by Ramakrishnan and Hadly 2009) including our own (Belle *et al.* 2006, Bramanti *et al.* 2009) and our close relatives (Belle *et al*. 2009, Fabre *et al.* 2009, see also Green *et al*. 2010).

The next section (6.4) provides an overview of some of the mathematical ideas underlying much of this work, before outlining the analysis techniques that allow for robust inference and parameter estimation. The final section (6.5) contains my original contributions to collaborative projects (Brace *et al.* (in prep), Bollongino *et al.* (under review)), using this approach to infer the past demography of two very different species.

## 6.4 The coalescent model

### 6.4.1 Derivation of the coalescent model

Recall that in the haploid Wright-Fisher model we have a population of effective size $N_e$. This population gives rise to, and is completely replaced by, a generation of offspring of the same size $N_e$. Reproduction is random, meaning mathematically that the number of offspring of each parent is determined by a random draw from a symmetric multinomial distribution. Over successive discrete generations this drift process results in the stochastic loss of lineages, but when viewed backwards through time it is directly equivalent to offspring individuals choosing their parent at random i.e. uniformly from the $N_e$ possible parents. So two

individuals in the current generation share a parent with probability $1/N_e$ and thus have distinct parents with probability $1 - 1/N_e$ (see Chapter 4; section 4.4.1). If we add a third individual there are now only $(N_e - 2)$ childless parents remaining, so the probability of this individual choosing a distinct parent is $(N_e - 2)/N_e = 1 - 2/N_e$. The total probability of all three having distinct parents is the product of these two terms, $(1 - 1/N_e) \times (1 - 2/N_e)$. Indeed, the probability of $n$ individuals having distinct parents in the previous generation $P(n)$ is:

$$P(n) = \prod_{i}^{n-1} \left( 1 - \frac{i}{N_e} \right)$$

Expanding this expression we get:

$$P(n) = 1 - \frac{1}{N_e} - \frac{2}{N_e} - \frac{3}{N_e} - \ldots - \frac{n-1}{N_e} + O(1/N_e^2)$$

where $O(1/N_e^2)$ gives expansion terms that tend to zero like $1/N_e^2$ (or faster) as $N_e$ tends to infinity. Therefore a good approximation for $P(n)$ when $N_e$ is large is:

$$P(n) \approx 1 - \frac{n(n-1)}{2N_e}$$

using the fact that the sum of the integers 1 to $n - 1$ is $n(n - 1)/2$. This can also be expressed as:

$$P(n) \approx 1 - \frac{\binom{n}{2}}{N_e}$$

where $\binom{n}{2}$ is the binomial coefficient, giving the number of possible combinations of two distinct elements from a set of size $n$. As the random choices of parents are independent across generations this probability $P(n)$ holds for *any* generation, so the probability of these $n$ individuals having distinct ancestors as far back as $t$ generations ago is $P(n)^t$. The probability that at least two of the $n$ individuals *share* a common ancestor $t + 1$ generations ago is then the product of $P(n)^t$ and the probability that all individuals in generation $t$ *do not* have distinct ancestor $[1 - P(n)]$. So the probability that the lineages of at least two individuals 'coalesce' in the $(t + 1)^{th}$ generation before present is $[1 - P(n)]P(n)^t$, which has the form of a geometric distribution i.e. $t$ 'failures' to coalesce followed by a 'success' in generation $t + 1$. This geometric distribution gives the waiting time to the first coalescent event amongst $n$ lineages, and can be approximated in continuous time by an exponential distribution with mean $1/(1 - P(n)) = N_e/\binom{n}{2}$:

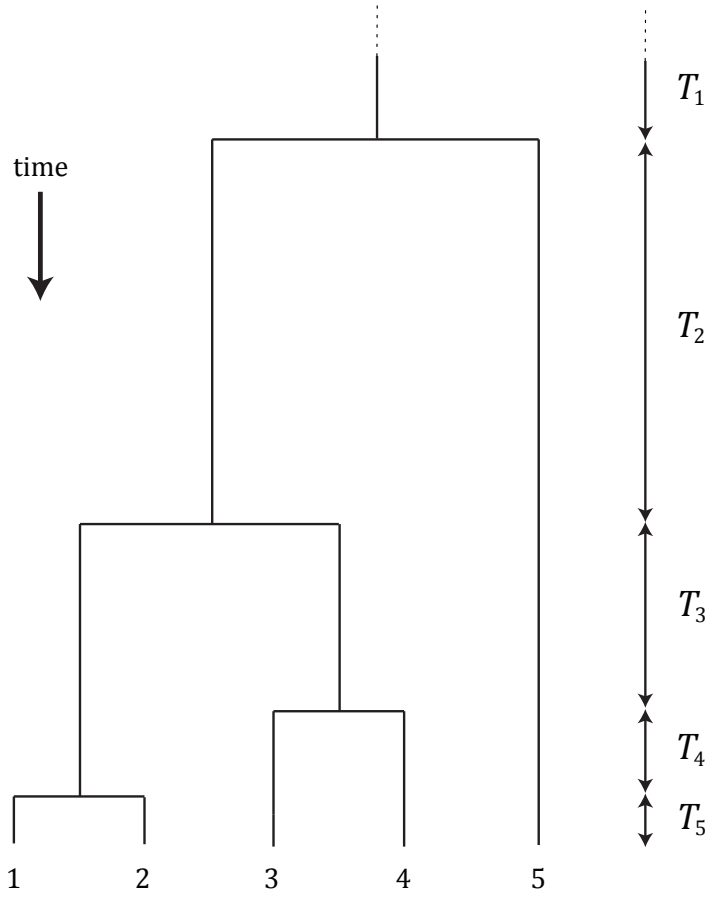
$$[1 - P(n)]P(n)^t \approx \frac{\binom{n}{2}}{N_e} e^{-\frac{\binom{n}{2}}{N_e}t}$$

As the lineages are unlabelled or 'exchangeable' this coalescence occurs between a randomly chosen pair of lineages. It is very important to note here that the coalescent approximation ignores the possibility of more than one coalescence per generation, which is reasonable with $N_e$ very large compared to sample size *n.* We are now left with $n - 1$ distinct lineages, and as all coalescent events are mutually independent the time to the next one is again exponentially distributed but with mean $N_e/\binom{n-1}{2}$ and, again, a random pair of lineages are chosen to coalesce.

In general, the time intervals $T_i$ in which there are $i$ distinct lineages are independently exponentially distributed with mean $N_e/\binom{i}{2}$, where $i = 2, 3, \ldots, n - 1$ (see figure 6.1). As we move back through time this process generates a stochastic bifurcating tree (it is bifurcating because multiple simultaneous coalescences are disallowed), with the internal nodes at the times $T_i$. The fact that the lineages are exchangeable and paired at random means that every possible tree topology is equally likely.

It is worth highlighting here that the coalescence times are dependent on the effective population size $N_e$, so that the time to a common ancestor will be (on average) greater in a larger population, as would be intuitively expected. This explicit link between the effective population size and the branch lengths of the sample genealogy underlies a lot of the power of this approach in inferring past demography from a relatively small number of samples.

**Figure 6.1**: An example of a coalescent genealogy for a sample of size 5 taken from the present generation. The $T_i$ are the independently exponentially distributed times between coalescent events, so give the durations of the interval in which *i* lineages are extant.




When we finally reach the $(n-1)^{th}$ coalescent event at the end of interval $T_2$, the remaining two distinct lineages coalesce, giving us the most recent common ancestor (MRCA) of all *n* sampled individuals. This approach allows us to easily

estimate the mean time to the MRCA ($T_{\mathrm{MRCA}}$) for these $n$ samples by simply summing the $n$ – 1 expected interval times $E(T_{\mathrm{i}})$:

$$E(T_{\mathrm{MRCA}}) = \sum_{\mathrm{i}=2}^{\mathrm{n}} E(T_{\mathrm{i}}) = \sum_{\mathrm{i}=2}^{\mathrm{n}} N_{\mathrm{e}} / \binom{i}{2} = 2N_{\mathrm{e}} \sum_{\mathrm{i}=2}^{\mathrm{n}} \frac{1}{i(i-1)}$$

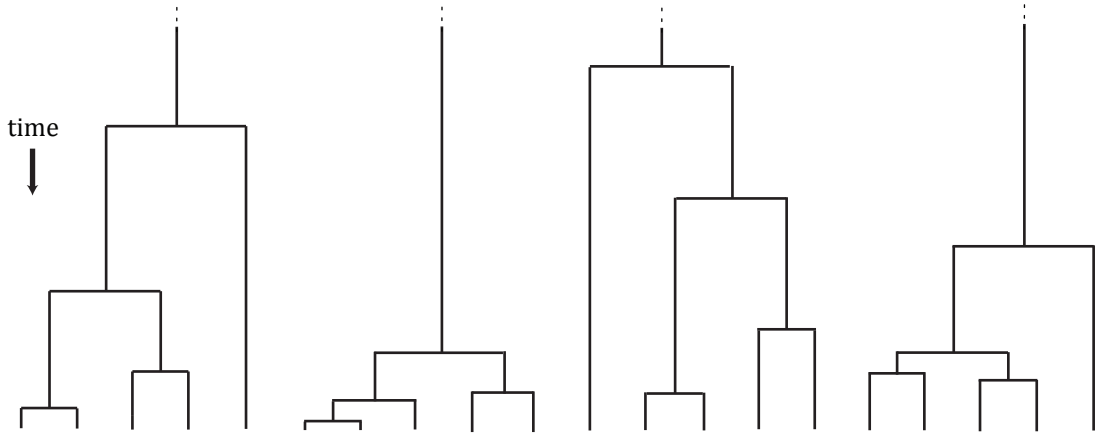
$$= 2N_{\mathrm{e}}\left(1 - \frac{1}{n}\right) \qquad (6.1)$$

It is interesting to note, from the expression on the far right of the first line, that the expected duration of interval $T_2$ (where there are just two remaining lineages) contributes more than 1/2 of the total time to the MRCA. This is intuitive, as for a fixed effective population size the lineages coalesce at rate $\binom{i}{2}/N_{\mathrm{e}}$, so coalescences take longer when there are fewer lineages and thus will tend to dominate the genealogy (see figures 6.1 and 6.2). The variance of $T_{\mathrm{MRCA}}$ is large compared to its mean, with the result that identically generated tree structures can be highly variable and, again, the final interval $T_2$ accounts for a large proportion of this variability (figure 6.2, see also Nordborg 2001).

Another thing to note from equation 6.1 is that as sample size $n$ increases towards the whole effective population size $N_{\mathrm{e}}$ the expected $T_{\mathrm{MRCA}}$ asymptotically approaches $2N_{\mathrm{e}}$ generations. This means that the entire population will share a common haploid i.e. Y-chromosome or mtDNA ancestor on average $\sim 2N_{\mathrm{e}}$ generations ago, recreating a result from classical population genetics (e.g. Wright 1931). It is also very easy to calculate the expected total length of the genealogy by simply summing

the product of the number of lineages $i$ and the expected duration of the interval $T_i$ i.e. $iE(T_i)$ over all intervals $i = 2, 3,..., n - 1$ (see for example Watterson 1975).

**Figure 6.2**: Four example coalescent genealogies for the same sample of size 5, showing the variability in branch lengths.




To summarize in practical terms, the $n$-coalescent is the set of $n - 1$ independent exponentially distributed coalescence times mapped onto the internal nodes of a random bifurcating tree with the $n$ sampled individuals at the tips. To reiterate, this is the probabilistic genealogy of just the $n$ individuals sampled in the present generation, so we do not concern ourselves with the vast number of lineages that existed in the population but were lost through drift before the sampled generation. This makes the coalescent approach extremely computationally efficient: to simulate a genealogy of $n$ samples, simply generate $n - 1$ independent exponential

variates (with appropriate means) and pair them with an independently created random bifurcating tree.

It should be noted that the original coalescent approximation is formally derived (Kingman 1982a, 1982b) with time scaled into coalescence units of $N$ generations and holds in the limit as $N$ tends to infinity. How this population size $N$ is related to the *effective* population size $N_e$ for a specific model of reproduction determines the time scaling factor one must use. Kingman (1982b) demonstrated that for a broad class of models described by an 'exchangeable' reproduction process (see Cannings 1974) – of which the Wright-Fisher model is simply a special case – the time-scaling is $N_e = N/\sigma^2$, where $\sigma^2$ is the variance in an individual's number of offspring. The distinctive birth-death process of the Moran model (see Chapter 5; section 5.4.2) means that it must be considered differently, but details are not given here (for example see Wakeley 2009; pages 66-68).

In the Wright-Fisher model reproduction occurs through a symmetric multinomial distribution, so that (when the population is large) the number of offspring for each individual $v_i$ has mean of 1 and variance $\sigma^2 = 1$. So the Wright-Fisher model has $N_e = N/1 = N$ and the informal derivation I have given above – which was intended to stress the importance of effective population size – is more than adequate for descriptive purposes. Wakeley (2009) provides a good recent review of the formal derivations and results for both the Wright-Fisher and Moran models (but also see Watterson 1975, Hudson 1983, Tajima 1983, Tavaré 1984, and reviews by Hudson 1991 and Donnelly and Tavaré 1995), as well as further discussion on appropriate time-scaling (Wakeley 2009; pages 68-74).

### 6.4.2 Extensions to the coalescent

Many important extensions to Kingman's *n*-coalescent have been made, including its expansion to include: recombination in diploid individuals, through use of the related, but reticulated, 'ancestral recombination graph' (Hudson 1983, 1991, Griffiths 1991); and selection, through the use of the 'ancestral selection graph' (Krone and Neuhauser 1997, Neuhauser and Krone 1997). Here, I briefly outline only those extensions that have direct relevance to the rest of this chapter (for reviews see Hudson 1991, Nordborg 2001). I do not give the mathematical details, but explain conceptually how these extensions work.

It has been shown that the requirement of a fixed population size can be relaxed to allow it to vary over time (see Griffiths and Tavaré 1994a, Donnelly and Tavaré 1995). This has the effect of scaling the set of coalescence times $T_i$, which in practical terms is just a matter of updating the effective population size $N_e(t)$ before the times are drawn from the exponential distribution with mean $N_e(t)/\binom{i}{2}$. However, in order for these models to remain tractable population size changes are often restricted to simple functions such as exponential growth or decline, or a piecewise combination of such functions (see for example Strimmer and Pybus 2001). Some population size changes leave a very obvious signature in the scaled branch lengths of a genealogy. One good example is the star-like phylogeny produced by an initially small but exponentially expanding population. The genealogy is dominated by very long recent branches, with the majority of coalescences occurring relatively rapidly in the smaller ancestral population.

Other authors have considered the effects of population substructure, with migration between a number of connected subpopulations or demes (see Hudson 1991, Donnelly and Tavaré 1995, Beerli and Felsenstein 1999, 2001, see also the 'island model' of Wright 1931). In this case the coalescent process can be thought of as the set of exponentially distributed waiting times to the next *even*t, whether that be a coalescence within a deme or the migration of a lineage to another deme, with relative probability governed by the inter-deme migration rate.

Finally – and most importantly for the integration of ancient DNA – Rodrigo and Felsenstein (1999) provided an extension to allow for observed samples taken from points other than the present day, called 'heterochronous' or 'serial' sampling. This was developed in order to look at rapidly evolving viruses – specifically HIV – where samples from multiple heterochronous generations were commonly obtained. The only major conceptual difference in this 'serial' model is that, contrary to the original coalescent model, the number of distinct lineages can *increase* as well as decrease as we move back in time incorporating the ancient samples (and their lineages) into the genealogy (for details see Rodrigo and Felsenstein 1999). By providing this direct evidence from the past, the inclusion of aDNA has been shown to improve inference on the timing and magnitude of past demographic events (see for example Drummond *et al.* 2003, Chan *et al.* 2006), and also allows the estimation of the mutation rate $\mu$, rather than the composite population parameter $\theta$.

### 6.4.3 Genetic modelling under the coalescent

I have so far deliberately avoided giving any genetic details in order to demonstrate that, under the assumptions of the coalescent, the structure of the genealogy of a sample is determined *solely* by demography, which can take into account both changing population size and substructure. However, the real power of the coalescent approach makes use of the fact (noted above) that the genetics of the individuals in the genealogy can be modelled separately and simply superimposed onto its branches. I have presented everything so far in terms of haploid individuals, but these coalescent genealogies hold for any non-recombining section of sequence, hence another common name 'gene genealogies' – which I will use hereafter.

We assume the population mutation parameter $\theta = 2N_e\mu$, where $\mu$ is the constant mutation rate per generation per sequence length or per locus, depending on the data type we want to simulate. For simulating DNA sequence data the infinite-sites model (Kimura 1969, Watterson 1975) is used, where every mutation is assumed to occur at a previously unmutated site. In this case $\mu$ is the overall rate per generation for the entire length of sequence, i.e. the rate per site multiplied by the number of base pairs (bp). For simulating a combination of discrete alleles or 'haplotypes', the infinite-alleles model (Kimura and Crow 1964) is used, where every mutation introduces a new allele into the population (see also Chapter 5; section 5.4.5.1). Microsatellite data can also be modelled as a simple stepwise random-walk process in which mutations either randomly increase or decrease copy number (see for example Ohta and Kimura 1973).

Under the coalescent assumptions, with time measured in coalescent units of $N_e$ generations, the occurrence of mutations can be described as a Poisson process with constant rate $\theta/2$ (see for example Donnelly and Tavaré 1995). This means that each branch of a gene genealogy will independently accumulate mutations in proportion to its length $\gamma$ (with mean number $\gamma\theta/2$). So starting at the root, or MRCA, of a gene genealogy simulated under a specified demographic model, we traverse the tree forward through time distributing mutations at rate $\theta/2$ on each branch, resulting in accumulated genetic diversity at the tips i.e. the modelled samples.

Many theoretical results can be derived on the expected distributions of assorted measures of genetic variation under this neutral model, which can then be used to test for deviations from neutrality or estimate population parameters for observed data (see Wakeley 2009 for an overview). But more importantly for our purposes here, both stochastic processes – the genealogical reconstruction and the mutation model – can be simulated exceptionally efficiently, so it is very easy to produce multiple simulated genetic datasets for a given demographic model with specified parameter values.

In generic terms, by comparing this simulated data – sampled from the appropriate time and deme in the model – with observed genetic data, we can make an evaluation of the 'quality' of the demographic model used. Rather than simply rejecting the demographic hypotheses that generate genetic variation that is inconsistent with what we observe, there exists a large and growing body of statistical techniques that allow for the explicit comparison of the competing demographic scenarios and the estimation of their underlying parameters. One such

technique, approximate Bayesian computation (ABC), was utilized in Chapter 2; section 2.6 and further detail is provided below.

An important result to note is that, perhaps contrary to intuition, increasing the sample size does not greatly increase the quality of inference. This is a natural consequence of the rate at which coalescences occur; with a large number of samples the lineages will quickly coalesce (i.e. very short branches) and all possible genealogies will always be dominated by the long deep branches that contain just a few remaining lineages (see Felsenstein 2006; figure 3). The number of mutations that accumulate down the lineages, and thus the genetic variation in the sample, depends on the *total* length of the genealogy, and increasing the number of samples has an increasingly negligible effect on this. Indeed, it has been shown that improving the estimation of population parameters is better served by sampling more *loci* rather than increasing the number of individuals sampled (Pluzhnikov and Donnelly 1996, Felsenstein 2006).

### 6.4.4 Statistical inference and model choice under the coalescent

Our aim of making inferences on population parameters, either through maximum likelihood or Bayesian approaches (see Beaumont and Rannala 2004 for a recent review), typically requires the calculation of likelihoods. This can be difficult if not impossible for anything more than the simplest of models (see Chapter 2; section 2.5.2).

Under the coalescent model, calculating the likelihood of the model parameter $\theta$ (= $2N_e\mu$) involves integration (or summation) over *every* possible genealogy. Direct computation is impossible as there are an astronomical number of different possible genealogies; with $n$ samples there are $n!(n-1)!/2^{n-1}$ possible tree topologies each of which has a distribution of the $n-1$ internal node times. One way round this is to recognize that a large proportion of the genealogies in the possible space have low probability given the data, i.e. low likelihood, and contribute only negligibly to the integral (or sum). Felsenstein and colleagues (Felsenstein 1992, Kuhner *at al.* 1995) developed Markov chain Monte Carlo (MCMC) methods that perform this summation by traversing just the non-negligible parts of genealogical space, allowing maximum likelihood estimation of various parameters (see also Griffiths and Tavaré 1994a, and Stephens and Donnelly 2000).

A Bayesian MCMC approach developed by Pybus, Rambaut, Drummond and colleagues (Pybus *et al.* 2000, Strimmer and Pybus 2001, Drummond *et al.* 2005) generates a 'Bayesian skyline plot', which is a posterior estimate of the effective size of a population back through time to the MRCA of the samples. This method has been implemented in the software package BEAST (Drummond and Rambaut 2007), which was briefly discussed in Chapter 3; section 3.6.1. One current drawback of this method, for our purposes at least, is that it considers lineages within only a single population, so is not ideally suited to inferring complex demographic histories (Anderson *et al.* 2005, Ramakrishnan and Hadly 2009) and can be sensitive to historical population structure if not properly accounted for (see Chapter 3; section 3.6.2, Powell *et al.* 2009).

However, likelihood-free approaches such as approximate Bayesian computation (ABC) allow us to avoid direct calculation or time-consuming approximation of likelihoods on the condition that we can generate large numbers of simulated datasets under a parameterized, and possibly complex, model (see Chapter 2; section 2.5.5). As has been discussed above, the efficiency of the coalescent approach makes this large-scale simulation straightforward and a wide range of software packages are now available for this purpose (see next section). A full description of parameter estimation under the ABC framework was provided in Chapter 2; section 2.6, so the only addition needed here is to detail its further application to evaluating competing demographic hypotheses i.e. model comparison.

This makes use of the idea of Bayes factors, which provide a summary of the evidence provided by the data in favour of a model from among all the possibilities considered. Recall Bayes theorem, which gives the posterior distribution of a model $H$ given the data $D$:

$$P(H \mid D) = \frac{P(D \mid H)P(H)}{P(D)}$$

where $P(D \mid H)$ is the probability of the data conditional on the model i.e. the likelihood, $P(H)$ is the prior probability of the model and $P(D)$ the normalizing constant. If we assume, for clarity, that we have just two competing models $H_1$ and $H_2$ (though the argument is easily extended to multiple models), then the Bayes factor is:

$$B_{12} = \frac{P(H_1 \mid D)}{P(H_2 \mid D)} = \frac{P(D \mid H_1)}{P(D \mid H_2)} . \frac{P(H_1)}{P(H_2)}$$

If the priors common to both models $P(H_i)$ are the same then this collapses to the ratio of the marginal likelihoods of each model:

$$B_{12} = \frac{L(H_1 \mid D)}{L(H_2 \mid D)}$$

This is a Bayesian analogue of the frequentist likelihood-ratio test, but rather than *maximizing* the likelihoods (see Chapter 2; section 2.5.1) it *averages* it over the model parameters, implicitly penalizing models with more parameters. This approach allows an explicit comparison of various competing demographic hypotheses with differing numbers of parameters, which can be highly complex, while guarding against over-parameterization or over-fitting of models.

Bayes factors can be empirically approximated by calculating the ratio of posterior probabilities of the competing models $P(H_i \mid S)$ given the summary statistics $S$ on the data $D$, through the rejection step of the ABC algorithm (see Chapter 2; section 2.5.5.3, and Pritchard *et al.* 1999). In practical terms, this means generating the same number of simulated datasets under each model, and rejecting for each model all but the proportion $F_\delta$ with the smallest Euclidean distances between simulated and observed statistics. The retained simulations from each model are then placed into a common pool and ordered according to renormalized Euclidean distance. The proportion of simulations due to each model within the top $n$-smallest Euclidean distances gives an estimate of the posterior probability of each model, and thus a measure of support given the evidence (see also Estoup *et al*. 2004 and Ray *et*

*al.* 2010). It should again be noted that there is still no formal method to decide on choice of summary statistics *S*, though see Chapter 2; section 2.5.5.7 for more on this.

### 6.4.5 Coalescent simulation software

An early algorithm / simulation package for generating simulated genetic data under the coalescent model *ms* was provided by Hudson (Hudson 1991 for the original algorithm, Hudson 2002). A recent study used this package to develop a complex serial-founder model of the expansion of humans from Africa, in order to explain the observed global patterns of human genetic variation (DeGiorgio *et al*. 2009). For more details of *ms* and other coalescent software packages such as: GeneTree (Griffiths and Tavaré 1994b, 1994c), IM (Nielsen and Wakeley 2001), LAMARC (v2.0 Kuhner 2006) and BEAST (Drummond and Rambaut 2007, also see Chapter 3; section 3.6.1), see the recent review by Kuhner (2008).

Another coalescent simulation package, SIMCOAL (Excoffier *et al.* 2000), allows the simulation of samples of DNA sequence, haplotype or microsatellite data under relatively complex demographic histories. This has since been updated as SIMCOAL2, which allows for multiple coalescences and the simulation of recombining genomic regions (Laval and Excoffier 2004). SIMCOAL has also been modified to include 'serial sampling' to allow application to both modern and ancient DNA in Serial SimCoal (Anderson *et al.* 2005). The same authors adapted this last package in order to allow for the use of Bayesian priors for the parameters in Bayesian Serial SimCoal (Anderson *et al.* 2005) – referred to as BayeSSC hereafter. Recall that under the Bayesian framework the focus is on making inferences on the unknown

parameters of a model given the data, where these parameters are given prior distributions that reflect our knowledge (or lack) about that parameter *before* the data has been considered (see Chapter 2; section 2.5.4). BayeSSC lends itself very neatly to direct integration with ABC, and all original work presented in the last section of this chapter utilizes this package coupled with my implementation, in the Python programming language (van Rossum *et al.* URL: http://www.python.org/), of ABC as canonically described by Beaumont *et al.* (2002).

### 6.4.6 Model-based and phylogeographic demographic inference

It is important to draw a distinction here between the coalescent approach and phylogeographic approaches to making demographic inferences from genetic data. Phylogeography uses the (apparent) geographical distribution of the clades of a gene genealogy or intra-specific phylogeny of a sample to infer population history. While the coalescent approach also relies on the idea of a gene genealogy, it does so within a population genetic framework, thus allowing for the explicit inclusion and subsequent evaluation of demographic processes.

The large and growing body of phylogeographic analyses (e.g. Avise *et al.* 1987, Cann *et al.* 1987, Richards *et al.* 1996, 2000, Templeton 1998, 2004) are generally motivated by empirical observation of an apparent geographic structuring of genetic variation. By creating a phylogeny of the observed haplotypes (i.e. the specific combinations of allelic states), the clades can be assigned to a spatial location based on the observed distribution of haplotype frequencies. Inferences are then made about past demographic processes that may have given rise to this geographic

distribution of clades or lineages, e.g. migration, gene flow, isolation by distance or range expansion. The inferred demographic history of a lineage is then interpreted as that of a population as a whole, ignoring the non-equivalence between lineages and populations. Under some simple conditions this equivalence may be justified, but in general there is a large potential for a subpopulation in a structured metapopulation and a specific lineage to become decoupled (i.e. 'lineage sorting', see Nielsen and Beaumont 2009; figure 2 and page 3).

Panchal and Beaumont (2007) provided theoretical evidence that even the most systematized of phylogeographic approaches, nested clade phylogeographic analysis (NCPA) (e.g. Templeton 1998, 2004) has a high false-positive rate when used in inference from a demographically explicit simulation model. They found that NCPA often inferred various demographic events from simulated data generated under a null panmictic model. They also showed that the relative frequencies of the type of event inferred closely matched those in the recent NCPA literature on real datasets (Panchal and Beaumont 2007; page 1475), with the implication that these inferences may be simply an artefact of the NCPA method itself.

In addition, the demographic phenomena of 'allele surfing' (as described in Chapter 1; section 1.3, and Chapter 2; section 2.3), which gives rise to previously unexpected spatial distributions of genetic variation, could have grave consequences for the validity of phylogeographic inference, which interprets spatial distributions of variation without the benefit of an underlying demographic model (Novembre and Rienzo 2009, see also Nielsen and Beaumont 2009, Ramakrishnan and Hadly 2009).

So while phylogeography privileges the network or phylogeny of haplotypes reconstructed from the data, the population genetics approach appropriately regards the phylogeny / gene genealogy as a random outcome of population level processes which is essentially uninteresting in itself, i.e. is a 'nuisance parameter'. This distinction is crucial in understanding the main conceptual difference between the two approaches, and is summarized neatly in the following passages from Nielsen and Beaumont (Nielsen and Beaumont 2009; pages 1 and 4):

> *"… while the tradition from phylogenetics is to estimate a tree and use the estimated tree to deduce evolutionary relationships, the population genetic tradition sees the tree as a random outcome of a population genetic process. Therein lies the fundamental difference between phylogeographic and theoretical population genetic thinking: phylogeographic studies traditionally assume that ancestral history can be directly deduced from estimated gene trees, whereas population genetic theory asserts that gene trees are random outcomes of stochastic population-level processes…"*

> *"There may be multiple demographic models that fit a gene tree equally well … Only by considering explicit demographic models will it be possible to determine which population histories are compatible with the data and which are not."*

The argument for model-based population genetic analysis is at present being made forcefully, and comprehensively, in the ongoing debate in *Molecular Ecology* (e.g. Pettit 2007, 2008, Beaumont and Panchal 2008, Beaumont *et al.* 2010, but see also Templeton 2004, 2008, 2009a, 2009b, 2010) and elsewhere (e.g. Panchal and Beaumont 2007, Knowles 2008, but see also Templeton 2009b).

### 6.4.7 Summary of the coalescent approach

In summary, the coalescent approach – as initially conceived by Kingman (1982a, 1982b), and extended (e.g. Griffiths and Tavaré 1994a, Donnelly and Tavaré 1995, Rodrigo and Felsenstein 1999) and implemented (e.g. Hudson 2002, Excoffier *et al.* 2000, Anderson *et al.* 2005) by numerous other authors – provides a powerful and exceptionally efficient framework for generating simulated genetic data under various, and increasingly complex, demographic histories.

This simulated data can then be used to make inferences about the demographic history of the population from which observed genetic data is obtained, using methods such as ABC (Tavaré *et al.* 1997, Pritchard *et al.* 1999, Beaumont *et al.* 2002).

## 6.5 Inferring demographic history: using the coalescent with ABC

Some successful examples of this coalescent / ABC approach include studies of the demographic history of species such as the cane toad (Estoup *et al.* 2001, Estoup *et al.* 2004), as well as human prehistory (Fagundes *et al.* 2007, Patin *et al.* 2009), including the initial colonisation of the Americas (Ray *et al.* 2010). Other studies, incorporating ancient DNA under a serial coalescent model, have investigated the past demography of a patagonian rodent species (Chan *et al.* 2006) and our own species: historical (Belle *et al.* 2006), prehistorical (Bramanti *et al.* 2009) and archaic (Belle *et al.* 2009), as well as Neandertals (Fabre *et al.* 2009).

The two following sections provide background and the details of my contribution to recent collaborative projects (Brace *et al.* (in prep), Bollongino *et al.* (under review)) on inferring the demographic history of two very different species using ancient, or historical, DNA and coalescent simulation coupled with ABC.

### 6.5.1 The demographic history of the Hispaniolan hutia[1]

#### 6.5.1.1 Background

The Caribbean island of Hispaniola, politically divided between Haiti in the west and the Dominican Republic in the east, has a complex geotechtonic history which underlies its interesting patterns of biodiversity. Hispaniola comprises two palaeo-islands, which came into contact around 10 million years ago (mya), though the division between them – the Neiba Valley depression – remained as a seaway until the Late Pleistocene (~125–10 kya). The north and south palaeo-islands have long been recognised as having distinct biota (Brace *et al.* (in prep)); and the south island contains a further biogeographic division. This Jacmel-Fauché depression, a sporadic sea passage in pre-Pleistocene times, also known as 'Bond's Line', bisects the south island (see figure 6.3).

The Hispaniolan hutia (*Plagiodontia aedium*) is one of only two surviving endemic land mammals on the island, previously thought to be extinct until the 1923 discovery of living specimens in Samaná Bay. Morphological differences between

[1] The content of this section appears in Brace *et al.* (in prep)

these individuals and zooarchaeological specimens led to suggestions of sympatric speciation, classifying the living hutia as the new species *Plagiodontia hylaeum* (Miller 1927). Following the 1947 discovery of a living hutia appearing to match the earlier *aedium* specimens, other phylogenetic and biogeographic hypotheses have been suggested. One proposes that there may be two subspecies; *P. aedium aedium* in Haiti and *P. aedium hylaeum* in the Dominican Republic (Anderson 1965). Another maintains that that all hutia populations across Hispaniola should be considered as conspecific (Woods and Howland 1979). Given this confusing taxonomic history of *Plagiodontia*, we sought to use the serial coalescent simulation / ABC inference approach, with both historical and modern genetic samples, to uncover more clearly its population history.

#### 6.5.1.2 Data

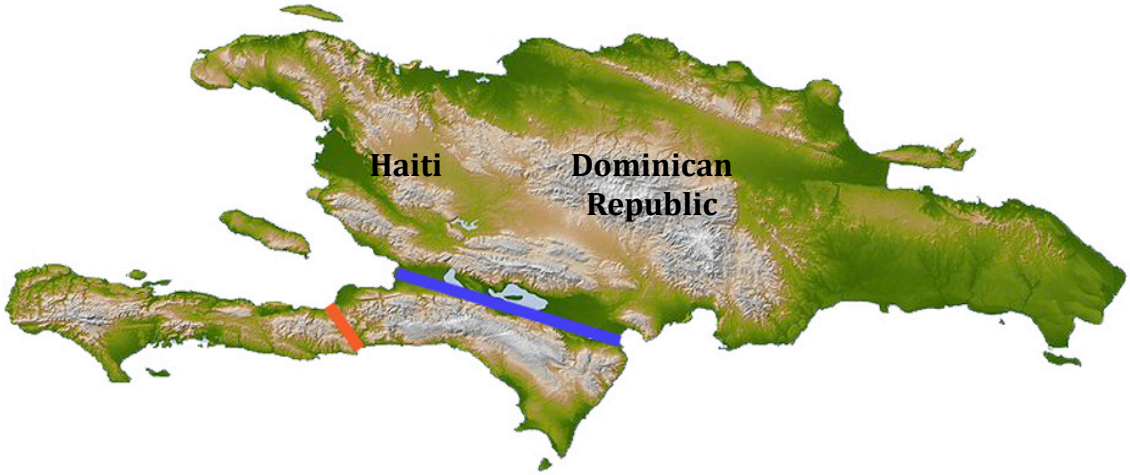
A total of 41 hutia samples were obtained through contemporary fieldwork and from early 20th century museum collections. These individuals provided 27 viable mtDNA sequences for the entire cytochrome b region, 1140 bp in length (for extraction details see Brace *et al.* (in prep)). 11 of these samples came from the north palaeo-island and 16 came from the south; of which 5 were from the southeast, and 11 from the southwest. Some initial summary statistics were calculated using Arlequin (Excoffier *et al.* 2005). Nucleotide diversity in the north island samples was around a tenth of that in the south (taken as a whole). Pairwise differences between the north and south were high compared to those across Bond's Line between the southeast and southwest (see Brace *et al.* (in prep)*;* tables 2 and 3). This, along with

evidence from other species, would suggest that Bond's Line is less of a barrier to migration than the divide between the north and south palaeo-islands; and the greater diversity in the south island further suggests a larger effective population size (given that it was colonized from the north, and thus cannot be older).

#### 6.5.1.3 Demographic models and ABC analysis

The geotectonic history of the island suggested two broad candidate models for investigating this species' demographic history. The first model ('North-South') assumed that the hutia colonised the south palaeo-island, from the north, soon after they joined ~10 mya, and that these two demes were connected by continual, but very low rate ($m_1$), migratory activity. The second model ('Bond's Line') resembled the first, but further assumes that the south island was separated, along Bond's Line, into southeast and southwest demes. The initial colonisation from the north populated the southeast deme, which in turn – after one generation – then colonized the southwest. Similar to the first model, the north and southeast demes were connected by migration at a low rate ($m_1$), but additionally the southeast and southwest were connected by migration at a potentially different rate ($m_2$), to reflect differences in the difficulty of crossing the two boundaries.

**Figure 6.3:** Map of Hispaniola – the blue line is the divide between north and south palaeo-islands, and the red line is Bond's Line.




In order to use BayeSSC (Anderson *et al.* 2005) we needed priors for all model parameters. The effective sizes $N_e$ of all demes in both models were given uniform priors on the range [1, 5 ×$10^5$]. The migration rate between the north and south/southeast demes $m_1$ was given a uniform prior on [0, 5 ×$10^{-5}$] for both models, while in the Bond's Line model the migration rate between southeast and southwest $m_2$ was given a uniform prior on [0, 5 ×$10^{-4}$], i.e. up to an order of magnitude larger.

Again using Arlequin (Excoffier *et al.* 2005) I calculated further summary statistics on the observed sequence data, which were partitioned into three sample groups based on their geographic origin: north, southeast and southwest. Within each sample group I calculated: the number of haplotypes, the number of segregating sites

and average pairwise difference; and between each group I calculated: $F_{ST}$, the number of haplotypes private to each sample, and average pairwise difference. This gave a total of 21 within- and between- sample group statistics. In the Bond's Line model I simulated three separate demes from which I took three simulated samples of matching size (i.e. 11 from the north deme, 5 from the southeast, and 11 from the southwest). In the North-South model I took one simulated sample from the north deme and then two (of size 5 and 11) from the single south deme. All simulated samples were taken from the 'present' generation in the model, as all observed samples were deemed sufficiently recent (20th century) and some samples came from contemporary fieldwork.

I performed 5 million coalescent simulations for each of these two models using BayeSSC (Anderson *et al.* 2005) which generates the simulated samples, as just described, by modelling a coalescent / mutation process with parameter values drawn from the prior distributions given above. I assumed an intergeneration time of 6 years (Nowak 1999, Kleiman *et al*. 1979), a mutation rate of 4% per million years – a reasonable estimate for rodent cytochrome b (Hadly *et al.* 2004) – and modeled a sequence 1140 bp in length, again in order to match the observed data. A transition/transversion ratio of 17.1:1 was assumed, while the mutation rate was Gamma-distributed with shape parameter 0.094. At the end of each simulation the same 21 within- and between-sample group summary statistics were calculated, and saved along with the parameter values that generated them.

I was interested in using ABC to compare these two models in order to better understand the demographic history of the hutia. This can be done through comparison of the posterior probability of each model given the observed data, as

outlined in section 6.4.4. Similarly to Chapter 2; section 2.6.1, I calculated the Euclidean distance between the simulated and observed summary statistics for each of the 5 million simulations under each model. Recall, this is simply the variance standardized distance between the vectors of simulated $s^{(i)}$ and observed statistics $s$, i.e. $||s^{(i)} - s||$. With tolerance proportion $F_\delta = 0.001$ I retained the 5,000 best-fitting parameter sets from each model, which were then pooled. These 10,000 simulations were then ordered by increasing Euclidean distance, which had been recalculated after normalization of the summary statistics using means and standard deviations computed on the pool. The relative proportion of simulations from each model in the $n$-smallest Euclidean distances then gave an estimate of the posterior probability of each model (see 6.4.4).

#### 6.5.1.4 Results

Following previous authors (Estoup *et al*. 2004, Ray *et al.* 2010), comparing the posterior probabilities at $n = 200$ showed that the Bond's Line model (0.77) was more likely than the simpler North-South model (0.23) given the observed summary statistics. Figure 6.4 demonstrates that this greater level of support for the Bond's Line model persisted over a large range for $n$.

Including the Bond's Line, with its two additional parameters, provided a better fit to the observed data than the simpler North-South model that ignored it. Parameter estimation for the Bond's Line model was carried out similarly to the ABC analysis previously described in Chapter 2; section 2.6.1. Briefly, with tolerance proportion $F_\delta = 0.001$ I retained the 5,000 best-fitting parameter sets produced by the

Bond's Line model, i.e. those with the smallest Euclidean distances between the observed and simulated summary statistics. The approximate joint posterior distribution was obtained by performing a local-linear multivariate weighted regression adjustment (see Beaumont *et al.* 2002 and Chapter 2; section 2.5.5.4), from which I derived marginal posterior distributions for each parameter (see figure 6.5).

**Figure 6.4:** Model posterior probabilities. Blue line: the two-deme North-South model. Red line: the three-deme Bond's Line model. At $n$ = 200 (vertical grey line): *Pr*(Bond's Line) = 0.77 and *Pr*(North-South) = 0.23.

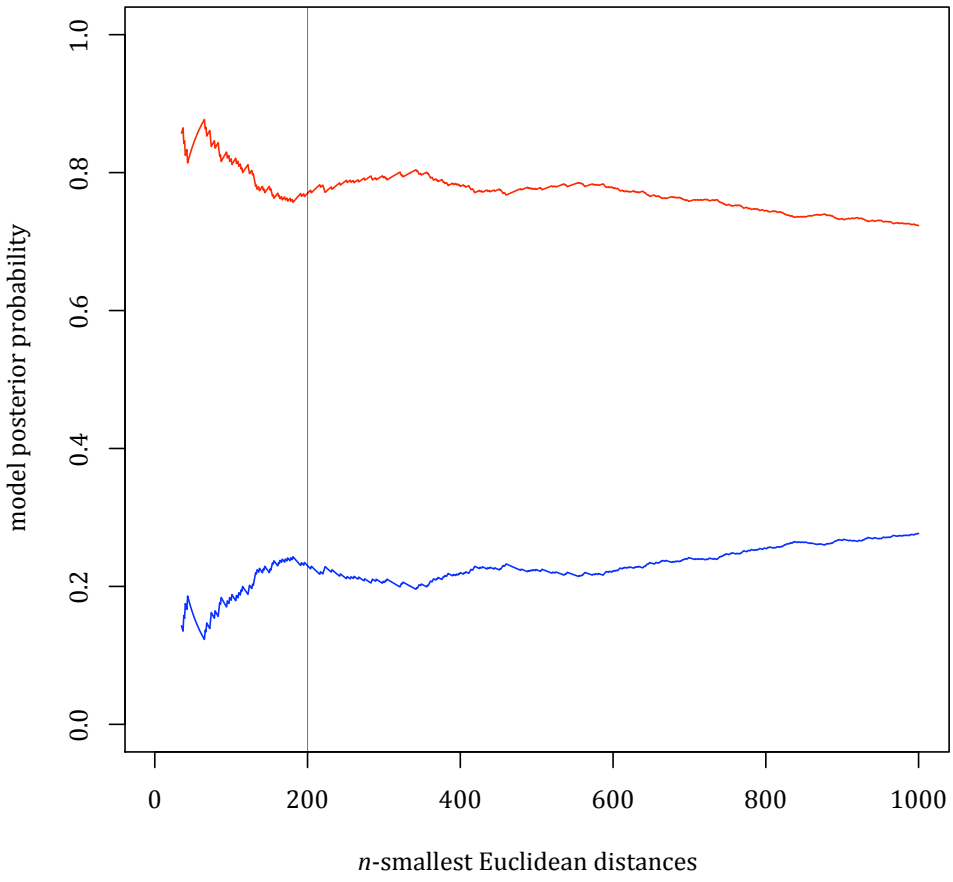


**Figure 6.5:** Marginal posterior densities of the five Bond's Line model parameters, each given over their prior range.

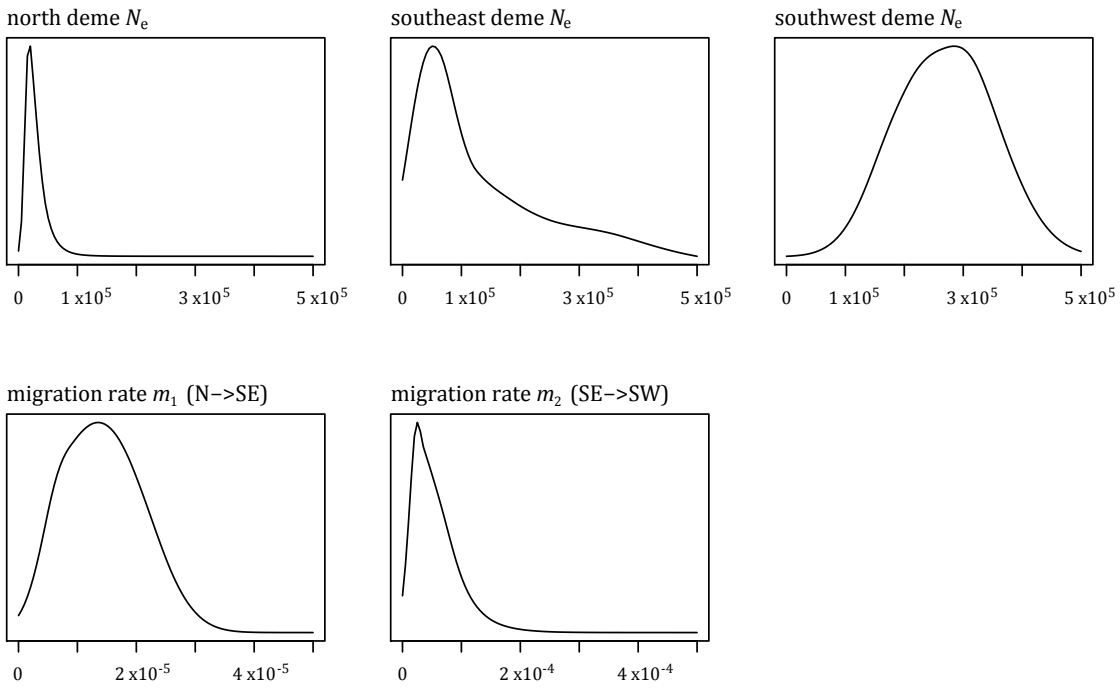



Table 6.1 gives the marginal posterior modes of each parameter, as well as 95% credible intervals (CIs). In contrast to a frequentist confidence interval an $x$% Bayesian credible interval contains the true parameter with probability $x$% (see Chapter 2; section 2.5.4).

**Table 6.1:** Marginal modal point estimates and 95% credible intervals (CI) for the parameters of the Bond's Line model.

| | **mode** | **95% CI** |
|---|---|---|
| north deme $N_e$ | 17300 | [7158, 75079] |
| southeast deme $N_e$ | 51866 | [7941, 456491] |
| southwest deme $N_e$ | 283968 | [110144, 437965] |
| migration rate $m_1$ (N->SE) | $1.35 \times 10^{-5}$ | $[2.65 \times 10^{-6}, 2.80 \times 10^{-5}]$ |
| migration rate $m_2$ (SE->SW) | $2.52 \times 10^{-5}$ | $[6.26 \times 10^{-6}, 1.57 \times 10^{-4}]$ |

#### 6.5.1.5 Conclusion

These posterior estimates indicate a much larger hutia population in the south palaeo-island – especially in the southwest – and that migration occurs across the Bond's Line at around twice the rate than across the north-south palaeo-island divide. The model choice procedure I employed demonstrated that the south palaeo-island has genetic structure, i.e. Bond's Line, that *needs* to be taken into account to better explain the contemporary variation.

As the first demographically explicit analysis of the hutia's population history, these results may prove important for the effective implementation of targeted conservation action. Given that very little of the endemic Caribbean island mammal fauna survived human colonizations – both the earliest Amerindian and the subsequent European colonizations (see for example Turvey 2009) – there is a pressing need to prevent further loss of mammalian diversity.

### 6.5.2 The domestication of Near Eastern taurine cattle[2]

#### 6.5.2.1 Background

The domestication of taurine cattle from wild aurochs began around 10,500 ya in southeastern Anatolia (Vigne 2008). Following an initial breeding phase lasting some 1,500 years in an area between the Levant, central Anatolia and western Iran, domestic cattle started to appear in southeastern Europe by 8,800 ya, southern Italy by 8,500 ya and central Europe by 8,000 ya (Vigne 2008). As previously described in Chapter 2; section 2.3, the archaeozoological (Vigne 2006, Vigne *et al.* 2007), residual lipid (Evershed *et al.* 2008) and lactase persistence (Itan *et al.* 2009) data point to an increasing economic importance for cattle as farming spread from Anatolia into central and southern Europe.

Consistent with the archaeozoology, the relatively low level of mitochondrial DNA (mtDNA) variation in modern European domestic cattle indicates a recent founding event (or events), and ancient mtDNA data suggests that there was little or no interbreeding with wild European aurochs (Edwards *et al.* 2007). However, the demographic history of Near Eastern cattle is difficult to infer directly from the number of surviving lineages in the modern population, due to stochasticity in the genealogical process (see section 6.4.6), as well as a lack of consensus on bovine mtDNA mutation rates (see for example Bradley *et al.* 1996, Edwards *et al*. 2007). By incorporating a set of ancient mtDNA sequences – recently obtained from specimens in or very near the original region of domestication – into a coalescent-based analysis

[2] The content of this section appears in Bollongino *et al.* (under review)

that allowed for the simultaneous estimation of the mutation rate, we aimed to more clearly uncover the history of Near Eastern cattle domestication.

#### 6.5.2.2 Data

The dataset comprises of 41 reliable mtDNA hypervariable region sequences: 15 from Neolithic to Iron Age Iranian domestic cattle (~8,000–1,900 ya) and 26 from modern Near Eastern cattle (see Bollongino *et al.* (under review)*;* supplementary information for details). These samples were split into 'ancient' and 'modern' sample groups. I used Arlequin (Excoffier *et al.* 2005) to calculate within each sample group: the number of haplotypes, haplotype diversity, the number of segregating sites, average pairwise difference and Tajima's D; and between the two groups: the number of private haplotypes, average pairwise difference, and $F_{ST}$. This gave a total of 14 within- and between-group summary statistics. Similar to the previous section, I used serial coalescent modelling as implemented by BayeSSC (Anderson *et al.* 2005) and ABC (Beaumont *et al.* 2002) to estimate the size of the domestication event and the mutation rate.

#### 6.5.2.3 Demographic model and ABC analysis

The model assumes an ancestral Near Eastern wild auroch female effective population size of 45,000 and an intergeneration time of 6 years (MacEachern *et al.* 2009). A single domestication event is presumed to occur 10,500 ya, in which a parameterized number of wild aurochs $N_e^{(d)}$ are sampled from the ancestral

population to form the population of domesticated cattle. While the domestication of wild aurochs is likely to have been a more complicated and lengthy process, I assumed a single instantaneous domestication event because in the absence of archaeological evidence for multiple events in the Near East (Vigne 2008), it is the most parsimonious model to first consider. The population of domesticates then grows exponentially to a modern Near Eastern female effective size of 800,000. The domestication event size parameter $N_e^{(d)}$ is given a uniform prior on the range [1, 5000], and the mutation rate $\mu$ is given a uniform prior over the interval 30.1% (Bradley *et al.* 1996) to 77.2% (Edwards *et al*. 2007) per million years.

I carried out 5 million coalescent simulations using BayeSSC (Anderson *et al.* 2005), and calculated the same 14 statistics for each simulated dataset, within and between the 'ancient' and 'modern' sample groups. With tolerance proportion $F_\delta$ = 0.001, I retained the 5,000 best-fitting parameter sets – tolerances of 0.002 and 0.005 gave very similar results. The approximate joint posterior distribution was obtained, as before, through local-linear weighted regression adjustment (Beaumont *et al.* 2002, see also Chapter 2; section 2.5.5.4), from which I derived marginal distributions for each parameter.

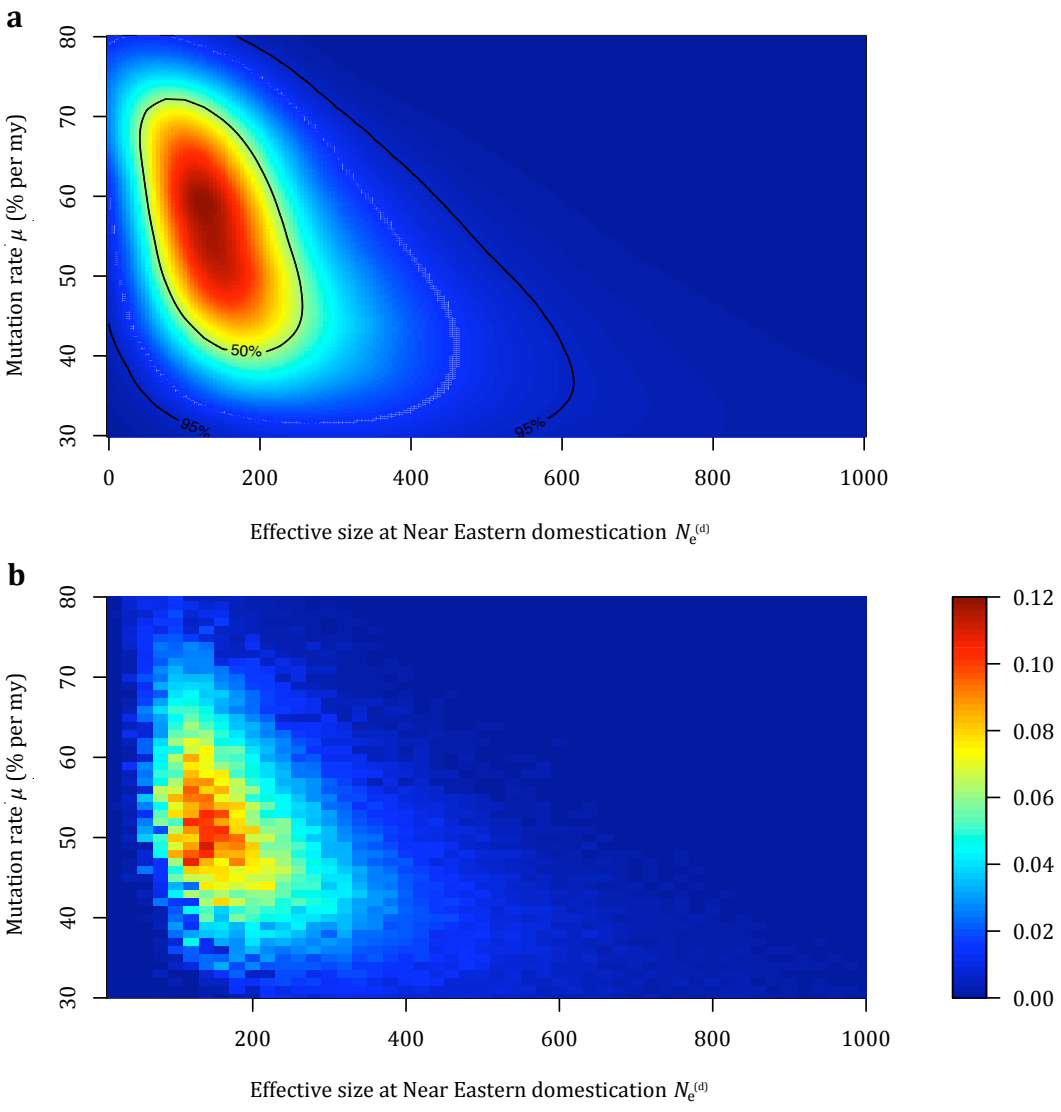
#### 6.5.2.4 Results

The marginal posterior mode for the effective female population size at domestication $N_e^{(d)}$ was 104 (95% CI: [46, 677]), and for the mutation rate $\mu$ was 49.9 % per million years (95% CI: [33.1, 74.9]). Figure 6.6a gives the joint posterior

distribution for $N_e^{(d)}$ and $\mu$, with the 50% and 95% CIs overlaid as contours. The joint posterior mode was at $N_e^{(d)} = 125$, $\mu = 59.1\%$. This small effective population size is consistent with similar estimates of pre-domestication ancestral $N_e$ across a range of modern taurine breeds based on linkage disequilibrium (Gibbs *et al.* 2009).

However, as mentioned above, the domestication process is likely to have been more complicated than a single event, so it could be the case that the model used may be mis-specified. In order to provide support for the model choice / parameter estimates, I employed an approach outlined by Voight *et al.* (Voight *et al.* 2005). This method, detailed in the next section, is based on combining simulated *p*-values from multiple summary statistics, and provides a criterion for rejecting demographic scenarios as incompatible with the observed values of the summary statistics.

**Figure 6.6: a** – the approximate joint posterior probability density of the effective female population size at domestication $N_e^{(d)}$ and the mutation rate per million years $\mu$ given the observed summary statistics. **b** – the confidence region for combinations of parameters $N_e^{(d)}$ and $\mu$ on a 51 × 51 grid created from $p$-values for the combined observed statistic $C_{obs}$.



#### 6.5.2.5 Model validation

Following Voight *et al.*, I first performed 10,000 coalescent simulations with Serial Simcoal (Anderson *et al.* 2005) on a 51 by 51 grid of values for parameters $N_e^{(d)}$ and $\mu$, covering the same range as the ABC analysis. For each parameter combination I used these simulations to estimate the probability $Pr_i$ of each simulated statistic being greater than each of the observed summary statistics. These probabilities were converted into two-tailed $p$-values $p_i$ for each statistic using $p_i = 1 - 2|0.5 - Pr_i|$. These 14 $p$-values were then used to calculate the combined $\chi^2$ statistic $C$ using Fisher's method:

$$C = -2\sum_{i=1}^{14} \ln(p_i)$$

This combined $C$ statistic calculated using the observed summary statistics is called $C_{obs}$. In order to create an expected distribution for $C$, I repeated this process but took each of the 10,000 simulations in turn and treated their generated summary statistics as I did the observed statistics previously. For each simulation I determined $p$-values for each statistic relative to the empirical distributions from the remaining 9,999 simulations. These $p$-values were combined as before and the resulting 10,000 values formed an empirical expected distribution for $C$, which was then used to calculate a two-tailed $p$-value for the observed statistic $C_{obs}$. This whole process was repeated for each parameter combination, resulting in a grid of $p$-values for $C_{obs}$ showing which combinations of $N_e^{(d)}$ and $\mu$ were (in)compatible with the observed

summary statistics i.e. a 'confidence region' (for another example see Voight *et al.* 2005; figure 2).

This confidence region, see figure 6.6b, was remarkably similar to the joint posterior density obtained through ABC (figure 6.6a), in that parameter combinations with higher posterior density also had high *p*-values. Indeed, the maximum *p*-value (= 0.111) was found at $N_e^{(d)} = 140$, $\mu = 49\%$ per million years, which was reasonably close to the joint mode estimated through ABC. This meant that we could *not* reject our model / parameter estimates given the observed summary statistics on the data, and we interpreted this as evidence that our model was reasonable.

#### 6.5.2.6 Conclusion

The low number of founding cattle we have inferred from this data may have major implications for Neolithic archaeology, as it indicates that cattle domestication was very likely initiated by restricted groups of humans. While we have used a simple demographic model, the previous section demonstrates that a single domestication event is consistent with the data, and in the absence of archaeological evidence to the contrary it is the most parsimonious explanation. If the first captive cattle were taken independently from a randomly mating wild population and interbred, then the effective population size estimate should still represent the number of Near Eastern aurochs that contributed to the domestic stock. If the first captive populations were isolated from one another then such structuring would preserve genetic variation and render the effective population size presented above as an overestimate.

While the small effective population size result is interesting in itself, it also provides another (indirect) source of evidence about the demography of the earliest Neolithic farmers. As mentioned in Chapter 2; section 2.7.1, the integration of multiple lines of evidence may provide further information on the human population dynamics of the Neolithic. Similar ongoing collaborative work on the domestication of other species such as goats and horses, will also potentially provide additional evidence on the demography of the earliest farming populations.

## 6.6 Chapter conclusion

The recent, and ongoing, development of coalescent theory as outlined in this chapter, has provided an efficient and powerful tool for generating simulated genetic data under increasingly complex demographic histories. In tandem with the rapidly developing field of approximate Bayesian computation (ABC), this approach enables robust inference on the demographic history of a population from observed genetic data – both modern and ancient. The two examples of my work given in the previous section demonstrate the efficacy of this approach; generating results which may aid conservation efforts (6.5.1.5), and shed new light on the processes of animal domestication at the beginning of the Neolithic (6.5.2.6).

# Concluding remarks

While I have drawn specific conclusions in each of the five previous chapters, there are a few points I wish to make in summary here. It is my hope that this thesis has shown clearly how demographic processes may affect variation in both genetic and cultural systems. Chapter 2 developed and utilized a geographically-explicit gene-culture coevolutionary model that incorporated the recently discovered demographic phenomenon of allele surfing, to explain the observed diversity across Europe of a recently derived lactase persistence allele. In Chapter 3, by adapting a previous model (Henrich 2004a) to allow for spatial substructure, I showed how heterogeneity in subpopulation density could account for heterogeneity in the accumulation of complex cultural traits. This result was used in tandem with genetic-based estimates of regional population density throughout the Late Pleistocene, to argue that demography was a critical factor in the heterogeneous appearance of the markers of modern human behaviour in the archaeological record; perhaps one of the most fundamental examples of cultural variation.

Chapters 4, 5 and 6 used existing and newly developed theory to explicitly demonstrate that observed genetic and cultural variation – from LBK pottery decoration styles to ancient cattle mtDNA – may be used to infer past population size and substructure and detect bottleneck events. This thesis considered both neutral (Chapters 4 and 5) and biased cultural transmission processes (Chapters 3 and 4), showing that there is room within the gene-culture framework for both approaches

and that they may provide complementary insights into different domains of human culture.

The gene-culture coevolution / cultural evolution framework allows the incorporation of the latest techniques and data from a wide range of previously disparate fields, and the work in Chapter 3 provides a good example of this. This work addressed a central and long-standing human evolutionary question by extending and integrating a mathematical model developed by a psychologist/anthropologist with evidence obtained from both population genetics and archaeology. Chapter 2 represents, to the best of my knowledge, the first application of the recently developed population genetic / statistical technique of ABC to both genetic *and* archaeological data. The integration of such evidence within the gene-culture coevolution framework enabled us to make inferences on key demographic model parameters, providing novel insights into another long-standing problem from our more recent evolutionary past.

It is also worth mentioning here the software package *Splatche* – and the imminent *Splatche 2* – developed by Mathias Currat, Nicolas Ray and Laurent Excoffier (2004). This package provides a framework for genetic coalescent modeling under a spatially- or geographically-explicit demographic history, and allows for spatial and temporal topographic and ecological heterogeneity. *Splatche* underlies many of the population genetic results reported in this thesis (e.g. Ray *et al.* 2005, Currat 2006), and was instrumental in understanding the allele surfing phenomenon (Klopfstein 2006, Currat *et al.* 2008). This is a powerful and flexible tool, and it will be interesting to see how this package can be integrated with archaeological, and perhaps linguistic, data in the future.

As outlined in the Introduction, the recent wide availability of high-resolution genomic data from populations around the globe has revolutionized the fields of population and statistical genetics. Novel and powerful statistical techniques have rapidly been developed in order to analyze and interpret this flood of new data. Two recent reviews from a gene-culture coevolutionary perspective incorporating these developments (Richerson *et al.* 2010, Laland *et al.* 2010) provide an excellent summary of the state of the field. Both sets of authors express optimism that the increasing ability to detect geographically structured selection pressures may unearth candidate loci requiring gene-cultural coevolutionary explanations. While challenges still remain, the potential for constructing an ever more detailed picture of our genetic and cultural evolutionary history is as immense as it is exciting.

# References


Aiello LC & Dunbar RIM (1993) Neocortex size, group size, and the evolution of language. *Current Anthropology* 34(2):184-193.

Akey JM, Zhang G, Zhang K, Jin L, & Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research* 12:1805-1814.

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA*, et al.* (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology* 2:e286.

Allen E, Beckwith J, Beckwith B, Chorover S, Culver D, & Duncan, M. *et al.* (13th November 1975) Against "Sociobiology". The New York Review of Books.

Ambrose SH (1998a) Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. *Journal of Human Evolution* 34(6):623-651.

Ambrose SH (1998b) Chronology of the Later Stone Age and food production in East Africa. *Journal of Archaeological Science* 25(4):377-392.

Ammerman AJ & Cavalli-Sforza LL (1984) *The Neolithic transition and the genetics of populations in Europe* (Princeton University Press, Princeton, NJ).

Anderson CNK, Ramakrishnan U, Chan YL, & Hadly EA (2005) Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics* 21(8):1733-1734.

Anderson S (1965) Conspecificity of Plagiodontia aedium and P. hylaeum (Rodentia). *Proceedings of the Biological Society of Washington* 78:95-98.

Aoki K (1986) A stochastic model of gene-culture coevolution suggested by the "culture historical hypothesis" for the evolution of adult lactose absorption in humans. *Proceedings of the National Academy of Sciences USA* 83:2929-2933.

Aoki K & Feldman MW (1987) Toward a theory for the evolution of cultural communication: coevolution of signal transmission and reception. *Proceedings of the National Academy of Sciences of the USA* 84(20):7164-7168.

Arbogast RM (1994) *Premiers élevages néolithqiues du Nord-Est de la France* (Ètudes et Rech. Arch. Univ. Liège, Liège).

Asch SE (1951) Effects of group pressure on the modification and distortion of judgments. *Groups, leadership and men*, ed Guetzkow H (Carnegie, Pittsburgh, PA), pp 177-190.

Atkinson QD, Gray RD, & Drummond AJ (2008) mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Molecular Biology and Evolution* 25(2):468-474.

Atran S (2001) The trouble with memes: inference versus imitation in cultural creation. *Human Nature* 12(4):351-381.

Aunger R (2002) *The electric meme: a new theory of how we think* (Free Press, New York).

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE*, et al.* (1987) Intraspecific phylogeography - the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18:489-522.

Axelrod R (1984) *The evolution of cooperation* (Basic, New York, NY).

Axtell RL (2001) Zipf distribution of U.S. firm sizes. *Science* 293:1818-1820.

Balasse M & Tresset A (2002) Early weaning of Neolithic domestic cattle (Bercy, France) revealed by intra-tooth variation in nitrogen isotope ratios. *Journal of Archaeological Science* 29:853-859.

Baldia MO (2003) Breaking unnatural barriers: Comparative archaeology, climate, and culture change in Central and Northern Europe (6000 - 2000 BC); in session "Comparative archaeology and paleoclimatology: sociocultural responses to a changing world" under the theme "Past human environments in modern contexts". *Fifth World Archaeology Congress, Washington DC, USA*.

Bánffy E (2004) *The 6th Millennium BC boundary in western Transdanubia and its role in the central European Neolithic transition (the Szentgyörgyvölgy-Pityerdomb settlement)* (Varia Arch. Hung., Budapest).

Bar-Yosef O (2002) The Upper Paleolithic Revolution. *Annual Review of Anthropology* 31:363-393.

Barabási A-L & Albert R (1999) Emergence of scaling in random networks. *Science* 286:509-512.

Barbujani G, Sokal RR, & Oden NL (1995) Indo-European origins: a computer-simulation test of five hypotheses. *American Journal of Physical Anthropology* 96(2):109-132.

Barkow JH, Cosmides L, & Tooby JE (1992) *The adapted mind: evolutionary psychology and the generation of culture* (Oxford University Press, Oxford).

Barnes I, Shapiro B, Lister A, Kuznetsova T, Sher A, Guthrie D, *et al.* (2007) Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Current Biology* 17:1072-1075.

Bartosiewicz L (2005) Animals, environment, and culture in the Neolithic of the Carpathian Basin and adjacent areas. *Unsettling the Neolithic*, eds Bailey D, Whittle A, & Cummings V (Oxbow), pp 51-63.

Bartosiewicz L (2007) Mammalian bone. *The early Neolithic on the Great Hungarian Plain: investigations of the Körös culture site of Ecsegfalva 23, County Békés*, ed Whittle A (Institute of Archaeology, Hungarian Academy of Sciences, Budapest).

Bayless TM, Paige DM, & Ferry GD (1971) Lactose intolerance and milk drinking habits. *Gastroenterology* 60:605-608.

Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics* 153:2013-2029.

Beaumont MA, Zhang W, & Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162(4):2025-2035.

Beaumont MA & Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics* 5(4):251-261.

Beaumont MA & Panchal M (2008) On the validity of nested clade phylogeographical analysis. *Molecular Ecology* 17:2563–2565.

Beaumont MA, Nielsen R, Robert C, Hey J, Gaggiotti O, Knowles L*, et al.* (2010) In defence of model-based inference in phylogeography. *Molecular Ecology* 19:436-446.

Beazley JD (1956) *Attic black-figure vase-painters* (Clarendon Press, Oxford).

Beerli P & Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763–773.

Beerli P & Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the USA* 98(8):4563-4568.

Beinhocker ED (2006) *The origin of wealth: evolution, complexity, and the radical remaking of economics* (Harvard Business School Press).

Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, Bertorelle G*, et al.* (2003) Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature Genetics* 35:311-313.

Belle EMS, Ramakrishnan U, Mountain JL, & Barbujani G (2006) Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. *Proceedings of the National Academy of Sciences USA*103(21):8012– 8017.

Belle EMS, Benazzo A, Ghirotto S, Colonna V, & Barbujani G (2009) Comparing models on the genealogical relationships among Neandertal, Cro-Magnoid and modern Europeans by serial coalescent simulations. *Heredity* 102:218-225.

Bellwood PS (2005) *The first farmers: the origins of agricultural societies* (Blackwell, Malden, MA).

Benecke N (1994a) *Der Mensch und seine Haustiere* (Theiss, Stuttgart).

Benecke N (1994b) *Archäozoologische Studien zur Entwicklung der Haustierhaltung in Mitteleuropa und Südskandinavien von den Anfängen bis zum ausgehenden Mittelalter. Schriften zur Ur- und Frühgeschichte* (Akademie Verlag Band 46, Berlin).

Benecke N (1998) Animal remains from the Neolithic and Bronze Age settlements at Kirklareli (Turkish Thrace). *Archaeozoology of the Near East III*, ed Buitenhuis H (ARC Publicaties, Groningen), pp 172-179.

Bentley RA & Shennan SJ (2003) Cultural transmission and stochastic network growth. *American Antiquity* 68:459-485.

Bentley RA, Hahn MW, & Shennan SJ (2004) Random drift and culture change. *Proceedings of the Royal Society B* 271:1443–1450.

Bentley RA, Lipo CP, Herzog HA, & Hahn MW (2007) Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior* 28:151-158.

Bentley RA, Ormerod P, & Batty M (2009) An evolutionary model of long tailed distributions in the social sciences. *arXiv* 0903.2533.

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA*, et al.* (2004) Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics* 74:1111-1120.

Bettinger RL & Eerkens J (1999) Typologies, cultural transmission, and the spread of bow-and-arrow technology in the prehistoric Great Basin. *American Antiquity* 64(2):231-242.

Bickerton D (1990) *Language and species* (University of Chicago Press, Chicago).

Birdsell JB, Bennett JW, Bicchieri MG, Claessen HJM, Gropper RC, Hart CWM*, et al.* (1973) A basic demographic unit. *Current Anthropology* 14(4):337-356.

Blackmore S (1999) *The meme machine* (Oxford University Press, Oxford).

Bloom G & Sherman P (2005) Dairying barriers affect the distribution of lactose malabsorption. *Evolution and Human Behavior* 26:301-333.

Blum MGB & François O (2009) Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing* 20(1):63-73.

Bocquet-Appel JP & Demars PY (2000) Neanderthal contraction and modern human colonization of Europe. *Antiquity* 74:544-552.

Boessneck JVD & Driesch A (1979) *Die Tierknochenfunde aus der Neolithischen Siedlung auf dem Fikirtepe bei Kadiköy am Marmarameer* (Munich, Inst. für Domestikationsforschung, Munich).

Bolender J (2007) Prehistoric cognition by description: a Russellian approach to the Upper Paleolithic. *Biology and Philosophy* 22:383-399.

Bollongino R, Edwards CJ, Alt KW, Burger J, & Bradley DG (2006) Early history of European domestic cattle as revealed by ancient DNA. *Biology Letters* 2:155–159.

Bollongino R, Burger J, Powell A, Mashkour M, Vigne J-D, & Thomas MG ((under review)) Modern Taurine cattle descended from small Near-Eastern herd.

Borgerhoff Mulder M, Nunn CL, & Towner MC (2006) Cultural macroevolution and the transmission of traits. *Evolutionary Anthropology* 15:52-64.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, & Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368(6470):455-457.

Bowles S (2004) *Microeconomics: behavior, institutions, and evolution* (Princeton University Press, Princeton, NJ).

Bowles S (2009) Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* 324(5932):1293-1298.

Boyd R & Richerson PJ (1985) *Culture and the evolutionary process* (University of Chicago Press, Chicago) pp viii, 331.

Boyd R & Richerson PJ (1995) Why does culture increase human adaptability? *Ethology and Sociobiology* 16:125-143.

Boyd R, Borgerhoff Mulder M, Durham WH, & Richerson PJ (1997) Are cultural phylogenies possible? *Human by nature: between biology and the social*

*sciences*, eds Weingart P, Mitchell SD, Richerson PJ, & Maasen S (Lawrence Erlbaum Associates, New Jersey), pp 355-386.

Boyd R & Richerson PJ (2002) Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology* 215:287–296.

Boyd R, Gintis H, Bowles S, & Richerson PJ (2003) The evolution of altruistic punishment. *Proceedings of the National Academy of Science USA* 100(6):3531–3535.

Boyd R & Richerson PJ (2009) Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B* 364:3281-3288.

Boyer P (1999) Cognitive tracks of cultural inheritance: how evolved intuitive ontology governs cultural transmission. *American Anthropologist* 100(4):876-889.

Brace S, Barnes I, Thomas MG, Powell A, Woolaver LG, & Turvey ST ((in prep)) Population history of the Hispaniolan hutia Plagiodontia aedium (Rodentia: Capromyidae): testing the model of ancient differentiation on a geotectonically complex Caribbean island.

Brainerd GW (1951) The place of chronological ordering in archaeological analysis. *American Antiquity* 16:301-313.

Bradley DG, MacHugh DE, Cunningham P, & Loftus RT (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences USA* 93(10):5131-5135.

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, *et al.* (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326:137-140.

Brantingham PJ, Krivoshapkin AI, Jinzeng L, & Tserendagva Y (2001) The initial Upper Paleolithic in Northeast Asia. *Current Anthropology* 42(5):735-747.

Brantingham PJ (2003) A neutral model of stone raw material procurement. *American Antiquity* 68(3):487-509.

Brantingham PJ, Kerry KW, Krivoshapkin AI, & Kuzmin YV (2004) Time-space dynamics in the early Upper Palaeolithic of Northeast Asia. *Entering America: Northeast Asia and Beringia before the last Glacial Maximum*, ed Madsen DB (University of Utah Press, Salt Lake City), pp 255-283.

Brantingham PJ (2007) A unified evolutionary model of archaeological style and function based on the Price equation. *American Antiquity* 72:395-416.

Brantingham PJ (2010) Detecting the simultaneous effects of selection and stochastic forces in archaeological assemblages. *Journal of Archaeological Science*.

Brockmann D, Hufnagel L, & Geisel T (2006) The scaling laws of human travel. *Nature* 439:462-465.

Brumm A & Moore MW (2005) Symbolic revolutions and the Australian archaeological record. *Cambridge Archaeological Journal* 15(2):157-175.

Burger J, Kirchner M, Bramanti B, Haak W, & Thomas MG (2007) Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences USA*104:3736-3741.

Buttenhuis H (1995) The faunal remains. *The Ilipinar Excavations I, Five Seasons of Fieldwork in NW Anatolia, 1987-91*, ed Roodenberg J (Nederlands Historisch-Archaeologisch Instituut, Istanbul), pp 151-156.

Campbell DT (1960) Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review* 67:380-400.

Campbell DT (1965) Variation and selective retention in socio-cultural evolution. *Social change in developing areas: a reinterpretation of evolutionary theory*, eds Barringer HR, Blanksten GI, & Mack RW (Schenkman, Cambridge, Mass), pp 19-49.

Cann RL, Stoneking M, & Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325(6099):31-36.

Cannings C (1974) The latent roots of certain Markov chains arising in genetics: a new approach. I. Haploid models. *Advances in Applied Probability.* 6:260-290.

Cavalli-Sforza LL & Feldman MW (1981) *Cultural transmission and evolution: a quantitative approach* (Princeton University Press, Princeton, N.J).

Cavalli-Sforza LL, Menozzi P, & Piazza A (1994) *The history and geography of human genes* (Princeton University Press, Princeton, NJ).

Cavalli-Sforza LL (2005) The Human Genome Diversity Project: past, present and future. *Nature Reviews Genetics* 6:333-340.

Chan YL, Anderson CNK, & Hadly EA (2006) Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genetics* 2(4):e59.

Chikhi L, Destro-Bisol G, Pascali V, Baravelli V, Dobosz M, & Barbujani G (1998) Clinal variation in the nuclear DNA of Europeans. *Human Biology* 70(4):643-657.

Chikhi L, Nichols RA, Barbujani G, & Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. *Proceedings of the National Academy of Sciences USA*99(17):11008-11013.

Childe VG (1942) *What happened in history* (Penguin Books).

Choi J-K & Bowles S (2007) The coevolution of parochial altruism and war. *Science* 318.

Chomsky N (2005) Three factors in language design. *Linguistic Enquiry* 36(1):1-22.

Clark JGD (1965) Radiocarbon dating and the expansion of farming culture from the Near East over Europe. *Proceedings of the Prehistoric Society* 31:57–73.

Clark PJ & Evans FC (1954) Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35(4):445-453.

Coelho M, Luiselli D, Bertorelle G, Lopes AI, Seixas S, Destro-Bisol G*, et al.* (2005) Microsatellite variation and evolution of human lactase persistence. *Human Genetics* 117:329-339.

Collard M, Shennan SJ, & Tehrani JJ (2006) Branching, blending, and the evolution of cultural similarities and differences among human populations. *Evolution and Human Behavior* 27(3):169-184.

Colledge S, Conolly J, & Shennan S (2004) Archaeobotanical evidence for the spread of farming in the Eastern Mediterranean. *Current Anthropology* 45:S35-S58

Conard N (2005) An overview of the patterns of behavioural change in Africa and Eurasia during the Middle and Late Pleistocene. *From Tools to Symbols from Early Hominids to Humans*, eds d'Errico F & Blackwell L (Wits University Press, Johannesburg), pp 294-332.

Cook W & Ormerod P (2003) Power law distribution of the frequency of demises of US firms. *Physica A* 324:207–212.

Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D*, et al.* (2009) The role of geography in human adaptation. *PLoS Genetics* 5(6):e1000500.

Copley MS, Berstan R, Dudd SN, Docherty G, Mukherjee AJ, Straker V*, et al.* (2003) Direct chemical evidence for widespread dairying in prehistoric Britain. *Proceedings of the National Academy of Sciences USA*100:1524-1529.

Copley MS, Berstan R, Mukherjee AJ, Dudd SN, Straker V, Payne S*, et al.* (2005) Dairying in antiquity III. Evidence from absorbed lipid residues dating to the British Neolithic. *Journal of Archaeological Science* 32:523-546.

Coultas J (2004) When in Rome... an evolutionary perspective on conformity. *Group Processes and Intergroup Relations* 7(4):317-331.

Craig OE, Chapman J, Heron C, Willis LH, Taylor G, Whittle A*, et al.* (2005) Did the first farmers of central and eastern Europe produce dairy foods? *Antiquity* 79:882–894.

Cribb RLD (1987) The logic of the herd: a computer simulation of archaeological herd structure. *Journal of Anthropological Archaeology* 6:376-415.

Crow JF & Kimura M (1970) *An introduction to population genetics theory* (Harper & Row, New York).

Currat M, Ray N, & Excoffier L (2004) Splatche: a program to simulate genetic diversity taking into account environmental heterogeneity, *Molecular Ecology Notes* 4(1): 139-142.

Currat M & Excoffier L (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biology* 2:e421.

Currat M & Excoffier L (2005) The effect of the Neolithic expansion on European molecular diversity. *Proceedings of the Royal Society B* 272(1564):679-688.

Currat M, Excoffier L, Maddison W, Otto SP, Ray N, Whitlock MC*, et al.* (2006) Comment on "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens" and "Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans". *Science* 313(5784):172

Currat M, Ruedi M, Petit RJ, & Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution* 62(8):1908-1920.

d'Errico F (2003) The invisible frontier: A multiple species model for the origin of behavioural modernity. *Evolutionary Anthropology* 12:188-202.

d'Errico F, Henshilwood C, Lawson G, Vanhaeren M, Tillier AM, Soressi M*, et al.* (2003) Archaeological evidence for the emergence of language, symbolism, and music: an alternative multidisciplinary perspective. *Journal of World Prehistory* 17:1-70

Darwin C (1859) *The origin of species* (Penguin, London) 1968 Ed.

Darwin C (1871) *The descent of man* (Oxford University Press, Oxford).

Dawkins R (1976) *The selfish gene* (Oxford University Press, Oxford).

Deacon TW (1997) *The symbolic species: the co-evolution of language and the brain* (W. W. Norton, New York, NY).

DeGiorgio M, Jakobsson M, & Rosenberg NA (2009) Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proceedings of the National Academy of Sciences USA*106(38):16057-16062.

Dennett D (1995) *Darwin's dangerous idea* (Simon & Schuster, New York, NY).

Diamond J (1998) *Guns, germs and steel* (Vintage, London).

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. *Nature* 418:700-707.

Dohnanyi JS (1969) Collisional model of asteroids and their debris. *Journal of Geophysical Research* 74(10):2531-2554.

Donnelly P & Tavaré S (1995) Coalescents and genealogical structure under neutrality. *Annual Review of Genetics* 29:401-421.

Donnelly P, Tavare S, Balding DJ, & Griffiths RC (1996) Estimating the age of the common ancestor of men from the ZFY intron. *Science* 272:1357-1359.

Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M*, et al.* (2004) Accelerated evolution of nervous system genes in the origin of Homo sapiens. *Cell* 119(7):1027-1040.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R, & Rodrigo AG (2003) Measurably evolving populations. *Trends in Ecology and Evolution* 18(9):481-488.

Drummond AJ, Rambaut A, Shapiro B, & Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* 22(5):1185-1192.

Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7:214.

Drummond AJ (2008) Reconstructing evolutionary bottlenecks using the coalescent. *http://bioinf.cs.auckland.ac.nz/index.php/2008/03/02/the-coalescent-for-bottlenecks/*.

Dunbar RIM (1993) Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences* 16(4):681-694.

Dunnell RC (1978) Style and function: a fundamental dichotomy. *American Antiquity* 43:192-202.

Dunnell RC (1980) Evolutionary theory and archaeology. *Advances in Archaeological Method and Theory* 3:35-99.

Dupanloup I, Bertorelle G, Chikhi L, & Barbujani G (2004) Estimating the impact of prehistoric admixture on the genome of Europeans. *Molecular Biology and Evolution* 21:1361-1372.

Durham WH (1991) *Coevolution: genes, culture, and human diversity* (Stanford University Press, Stanford).

Durham WH (1992) Applications of evolutionary culture theory. *Annual Review of Anthropology* 21:331-355.

Edmonds CA, Lillie AS, & Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences USA* 101(4):975-979.

Edwards AWF (2003) Human genetic diversity: Lewontin's fallacy. *BioEssays* 25:798-801.

Edwards CJ, Bollongino R, Scheu A, Chamberlain A, Tresset A, Vigne JD*, et al.* (2007) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society B* 274(1616):1377-1385.

Eerkens JW & Lipo CP (2005) Cultural transmission, copying errors, and the generation of variation in material culture and the archaeological record. *Journal of Anthropological Archaeology* 24:316-334.

Eerkens J, Bettinger RL, & McElreath R (2006) Cultural transmission, phylogenetics and the archaeological record. *Mapping Our Ancestors: Phylogenetic Approaches in Anthropology and Prehistory*, eds Lipo CP, O'Brien MJ, Collard M, & Shennan SJ (AldineTransaction, New Brunswick, NJ), pp 169–184.

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T*, et al.* (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418(6900):869-872.

Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, & Jarvela I (2002) Identification of a variant associated with adult-type hypolactasia. *Nature Genetics* 30(2):233-237.

Enattah NS, Trudeau A, Pimenoff V, Maiuri L, Auricchio S, Greco L*, et al.* (2007) Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. *American Journal of Human Genetics* 81:615-625.

Enattah NS, Jensen TG, Nielsen M, Lewinski R, Kuokkanen M, Rasinpera H*, et al.* (2008) Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *American Journal of Human Genetics* 82:57-72.

Enquist M & Ghirlanda S (2007) Evolution of social learning does not explain the origin of human cumulative culture. *Journal of Theoretical Biology* 246:129–135.

Enquist M, Ghirlanda S, Jarrick A, & Wachtmeister C-A (2008) Why does human culture increase exponentially? *Theoretical Population Biology* 74:46–55.

Erdős P & Rényi A (1960) On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci* 5:17-61.

Estoup A, Wilson IJ, Sullivan C, Cornuet J-M, & Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, 'Bufo marinus' *Genetics* 159:1671-1687.

Estoup A, Beaumont M, Sennedot F, Moritz C, & Cornuet JM (2004) Genetic analysis of complex demographic scenarios: Spatially expanding populations of the cane toad, Bufo marinus. *Evolution* 58(9):2021-2036.

Evans J (1850) On the date of British coins. *The Numismatic Chronicle and Journal of the Numismatic Society* 12:127-137.

Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM*, et al.* (2005) Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* 309(5741):1717-1720.

Evans PD, Mekel-Bobrov N, Vallender EJ, Hudson RR, & Lahn BT (2006) Evidence that the adaptive allele of the brain size gene microcephalin introgressed into Homo sapiens from an archaic Homo lineage. *Proceedings of the National Academy of Sciences USA* 103(48):18178-18183.

Everett D (2005) Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language. *Current Anthropology* 46(4):621-634.

Evershed RP, Payne S, Sherratt AG, Copley MS, Coolidge J, Urem-Kotsu D*, et al.* (2008) Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* 455:528-531.

Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* 3(1):87-112.

Ewens WJ (2004) *Mathematical population genetics* (Springer-Verlag, New York, NY) 2nd Ed.

Excoffier L, Novembre J, & Schneider S (2000) SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity* 91(6):506-509.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology* 13(4):853-864.

Excoffier L, Laval G, & Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* 1:47-50.

Fabre V, Condemi S, & Degioanni A (2009) Genetic evidence of geographical groups among Neanderthals. *PLoS One* 4(4):e5151.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL*, et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences USA*104(45):17614-17619.

Feldman MW & Cavalli-Sforza LL (1976) Cultural and biological evolutionary processes, selection for a trait under complex transmission. *Theoretical Population Biology* 9:238-259.

Feldman MW & Laland KN (1996) Gene-culture coevolutionary theory. *Trends Ecol Evol* 11:453-457.

Felsenstein J (1971b) The rate of loss of multiple alleles in finite haploid populations. *Theoretical Population Biology* 2:391-403.

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics* 22:521-565.

Felsenstein J (1992) Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genetical Research* 59:139-147.

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* 23(3):691-700.

Fisher RA (1930 ) *The genetical theory of natural selection* (Clarendon Press, Oxford).

Flatz G & Rotthauwe HW (1973) Lactose nutrition and natural selection. *Lancet* 2:76-77.

François O, Currat M, Ray N, Han E, Excoffier L, & Novembre J (2010) Principal component analysis under population genetic models of range expansion and admixture. *Molecular Biology and Evolution* 27(6):1257-1268.

Frank SA (1995) George Price's contributions to evolutionary genetics. *Journal of Theoretical Biology* 175:373–388.

Frank SA (2009) The common patterns of nature. *Journal of Evolutionary Biology* 22:1563–1585.

Fu Y-X & Li W-H (1997) Estimating the age of the common ancestor of a sample of DNA sequences. *Molecular Biology and Evolution* 14(2):195-199.

Gabaix X (1999) Zipf's Law for cities: an explanation. *Quarterly Journal of Economics* 114:739–767.

Ghirlanda S & Enquist M (2007) Cumulative culture and explosive demographic transitions. *Quality and Quantity* 41(4):591-600.

Ghirlanda S, Enquist M, & Perc M (2010) Sustainability of culture-driven population dynamics. *Theoretical Population Biology* 77:181-188.

Gibbs RA, Taylor JF, Van Tassell CP, & The Bovine HapMap Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324(528-532).

Gilbert MTP, Bandelt H, Hofreiter M, & Barnes I (2005) Assessing ancient DNA studies. *Trends in Ecology and Evolution* 20:541-544.

Gilbert SL, Dobyns WB, & Lahn BT (2005) Genetic links between brain development and brain evolution. *Nature Reviews Genetics* 6(7):581-590.

Gintis H (2007) A framework for the unification of the behavioral sciences. *Behavioral and Brain Sciences* 30:1-61.

Gintis H (2009) *Game theory evolving* (Princeton University Press, Princeton, NJ) 2nd Ed.

Goebel T, Waters MR, & O'Rourke DH (2008) The Late Pleistocene dispersal of modern humans in the Americas. *Science* 319(5869):1497-1502.

Gray RD & Jordan FM (2000) Language trees support the express-train sequence of Austronesian expansion. *Nature* 405:1052-1055.

Gray RD & Atkinson QD (2003) Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426:435-439.

Gray RD, Drummond AJ, & Greenhill SJ (2009) Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323:479-483.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, *et al.* (2010) A draft sequence of the Neandertal genome. *Science* 328(5979):710-722.

Greenberg JH, Turner II CG, & Zegura SL (1986) The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence. *Current Anthropology* 27:477-497.

Greenhill SJ, Currie TE, & Gray RD (2009) Does horizontal transmission invalidate cultural phylogenies? *Proceedings of the Royal Society B* 276:2299-2306.

Griffiths RC (1980) Lines of descent in the diffusion approximation of neutral Wright-Fisher models. *Theoretical Population Biology* 17:37-50.

Griffiths RC (1991) The two-locus ancestral graph. *Selected Proceedings of the Symposium on Applied Probability, Sheffield, 1989. IMS Lecture Notes - Monograph Series*, eds Basawa IV & Taylor RL (Institute of Mathematical Statistics, Hayward, California), Vol 18, pp 100-117.

Griffiths RC & Tavaré S (1994a) Ancestral inference in population genetics. *Statistical Science* 9:307-319.

Griffiths RC & Tavaré S (1994b) Simulating probability distributions in the coalescent. *Theoretical Population Biology* 46:131-159.

Griffiths RC & Tavaré S (1994c) Sampling theory for neutral alleles in a varying environment. *Philosophical Transaction of the Royal Society B* 344:403-410.

Gutenberg B & Richter CF (1944) Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* 34:185-188.

Haberle SG & David B (2004) Climates of change: human dimensions of Holocene environmental change in low latitudes of the PEPII transect. *Quaternary International* 118-119:165-179.

Hadly EA, Ramakrishnan U, Chan YL, van Tuinen M, O'Keefe K, Spaeth PA*, et al.* (2004) Genetic response to climatic change: insights from ancient DNA and phylochronology. *PLoS Biology* 2(10):e290.

Hahn MW & Bentley RA (2003) Drift as a mechanism for cultural change: an example from baby names. *Proceedings of the Royal Society B* 270:S1-S4.

Hallatschek O & Nelson DR (2009) Life at the front of an expanding population. *Evolution* 64(1):193-206.

Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, & Excoffier L (2005) Bayesian estimation of recent migration rates after a spatial expansion. *Genetics* 170(1):409-417.

Hamilton WD (1970) Selfish and spiteful behaviour in an evolutionary model. *Nature* 228:1218-1220.

Hamilton WD (1975) Innate social aptitudes of man: an approach from evolutionary genetics. *Biosocial Anthropology*, ed Fox RG (Wiley, New York), pp 133-155.

Handley LJ, Manica A, Goudet J, & Balloux F (2007) Going the distance: human population genetics in a clinal world. *Trends in Genetics* 23(9):432-439.

Harpending H & Rogers A (2000) Genetic perspectives on human origins and differentiation. *Annual Review of Genomics and Human Genetics* 1:361-385

Harpending HC, Sherry ST, Rogers AR, & Stoneking M (1993) The genetic structure of ancient human populations. *Current Anthropology* 34:483-496.

Hassan FA (1981) *Demographic archaeology* (Academic Press, New York).

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications *Biometrika* 57(1):97-109.

Hawks J, Wang ET, Cochran GM, Harpending HC, & Moyzis RK (2007) Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences USA* 104(52):20753-20758.

Henrich J & Boyd R (1998) The evolution of conformist transmission and between-group differences. *Evolution and Human Behavior* 19:215-242.

Henrich J (2001) Cultural transmission and the diffusion of innovations: adoption dynamics indicate that biased cultural transmission is the predominant force in behavioral change. *American Anthropologist* 103(4):992-1013.

Henrich J & Gil-White FJ (2001) The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior* 22(3):165-196.

Henrich J & Boyd R (2002) On modeling cognition and culture: why cultural evolution does not require replication of representations. *Journal of Cognition and Culture* 2(2):87-112.

Henrich J & McElreath R (2003) The evolution of cultural evolution. *Evolutionary Anthropology* 12(3):123-135.

Henrich J (2004a) Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses – the Tasmanian case. *American Antiquity* 69:197-214.

Henrich J (2004b) Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization* 53:3-35.

Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H*, et al.* (2005) 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences* 28:795-855.

Henrich J (2006) Understanding cultural evolutionary models: a reply to Read's critique. *American Antiquity* 71(4):771-782.

Henrich J & McElreath R (2008) Dual inheritance theory: the evolution of human cultural capacities and cultural evolution. *Oxford Handbook of Evolutionary Psychology*, eds Dunbar R & Barrett L (Oxford University Press, Oxford), pp 555-570.

Henshilwood CS, d'Errico F, Yates R, Jacobs Z, Tribolo C, Duller GAT*, et al.* (2002) Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science* 295:1278–1280.

Henshilwood CS & Marean CW (2003) The origins of modern human behavior: critique of the models and their test implications. *Current Anthropology* 44(5):627-651.

Henshilwood CS, d'Errico F, Vanhaeren M, van Niekerk K, & Jacobs Z (2004) Middle Stone Age shell beads from South Africa. *Science* 304(5669):404.

Herzog HA, Bentley RA, & Hahn MW (2004) Random drift and large shifts in popularity of dog breeds. *Proceedings of the Royal Society B* 271 Suppl 5:S353-356.

Hofer T, Ray N, Wegmann D, & Excoffier L (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Annals of Human Genetics* 73(1):95-108.

Hoffecker JF (2005) Innovation and technological knowledge in the Upper Paleolithic of Northern Eurasia. *Evolutionary Anthropology* 14:186-198.

Hofreiter M, Serre D, Poinar HN, Kuch M, & Paabo S (2001) Ancient DNA. *Nature Reviews Genetics* 2:353-359.

Holden CJ (2002) Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum parsimony analysis. *Proceedings of the Royal Society B* 269(1493):793-799.

Holden CJ & Mace R (1997) Phylogenetic analysis of the evolution of lactose digestion in adults. *Human Biology* 69:605-628.

Holden CJ & Mace R (2003) Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proceedings of the Royal Society B* 270:2425-2433.

Hollox EJ, Poulter M, Zvarik M, Ferak V, .Krause A, Jenkins T*, et al.* (2001) Lactase haplotype diversity in the Old World. *American Journal of Human Genetics* 68:160-172.

Howells WW (1976) Explaining modern man: evolutionists versus migrationists. *Journal of Human Evolution* 5(5):477-495.

Hudjashov G, Kivisild T, Underhill PA, Endicott P, Sanchez JJ, Lin AA*, et al.* (2007) Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis. *Proceedings of the National Academy of Sciences USA* 104(21):8726-8730.

Hudson RR (1983) Testing the constant rate neutral allele model with protein sequence data. *Evolution* 37:203–217.

Hudson RR (1991) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, eds Futuyma D & Antonovics J (Oxford University Press, Oxford), Vol 7, pp 1-44.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model. *Bioinformatics* 18:337-338.

Ingman M, Kaessmann H, Pääbo S, & Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408(6813):708-713.

Ingram CJ, Elamin MF, Mulcare CA, Weale ME, Tarekegn A, Raga TO*, et al.* (2007) A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Human Genetics* 120:779-788.

Ingram CJ, Mulcare CA, Itan Y, Thomas MG, & Swallow DM (2009) Lactose digestion and the evolutionary genetics of lactase persistence. *Human Genetics* 124:579-591.

Itan Y, Powell A, Beaumont MA, Burger J, & Thomas MG (2009) The origins of lactase persistence in Europe. *PLoS Computational Biology* 5:e1000491.

Itan Y, Jones BL, Ingram CJ, Swallow DM, & Thomas MG (2010) A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evolutionary Biology* 10:36.

Jablonski NG & Chaplin G (2000) The evolution of human skin coloration. *Journal of Human Evolution* 39(1):57-106

Jacobs Z, Roberts RG, Galbraith RF, Deacon HJ, Grün R, Mackay A*, et al.* (2008) Ages for the Middle Stone Age of Southern Africa: implications for human behaviour and dispersal. *Science* 322:733-735.

James HVA & Petraglia MD (2005) Modern human origins and the evolution of behavior in the Later Pleistocene record of South Asia. *Current Anthropology* 46, Supplement, Dec., 2005:3-27.

Jeong H, Tombor B, Albert R, Oltvai ZN, & Barabási A-L (2000) The large-scale organization of metabolic networks. *Nature* 407:651-654.

Jorde LB, Watkins WS, bamshad MJ, Dixon ME, Ricker CE, Seielstad MT*, et al.* (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *American Journal of Human Genetics* 66(3):979-988.

Joyce P & Marjoram P (2008) Approximately sufficient statistics and Bayesian computation. *Statistical Applications in Genetics and Molecular Biology* 7:26.

Kaessmann H, Wiebe V, Weiss G, & Pääbo S (2001) Great ape DNA sequences reveal a reduced diversity and expansion in humans. *Nature Genetics* 27:155-156.

Kandler A (2009) Demography and language competition. *Human Biology* 81(2-3):181–210.

Kandler A & Laland KN (2009) An investigation of the relationship between innovation and cultural diversity. *Theoretical Population Biology* 76:59-67.

Kandler A & Steele J (2010) Social learning, economic inequality, and innovation diffusion. *Innovation in cultural systems: contributions from evolutionary anthropology*, eds O'Brien M & Shennan SJ (MIT Press, Cambridge, MA).

Kimura M & Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49:725-738.

Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61:893-903.

Kimura M (1983) *The neutral allele theory of molecular evolution* (Cambridge University Press, Cambridge).

Kingman JFC (1982a) The coalescent. *Stochastic Processes and their Applications* 13:235-248.

Kingman JFC (1982b) On the genealogy of large populations. *Journal of Applied Probability* 19:27-43.

Kivisild T (*personal communication,* 2009).

Kleiman DG, Eisenberg JF & Maliniak E (1979) Reproductive parameters and productivity in caviomorph rodents. *Vertebrate ecology in the northern neotropics*, ed Eisenberg JF (Smithsonian Institution Press, Washington, D.C.), pp 173-183.

Klein RG (1999) *The human career: human biological and cultural origins* (The University of Chicago Press, Chicago) 2nd Ed.

Klein RG (2000) Archaeology and the evolution of human behavior. *Evolutionary Anthropology* 9(1):17-36.

Kline MA & Boyd R (2010) Population size predicts technological complexity in Oceania. *Proceedings of the Royal Society B* 277:2559-2564.

Klopfstein S, Currat M, & Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution* 23(3):482-490.

Knowles LL (2008) Why does a method that fails continue to be used? *Evolution* 62:2713-2717.

Kohler TA, VanBuskirk S, & Ruscavage-Barz S (2004) Vessels and villages: evidence for conformist transmission in early village aggregations on the Pajarito Plateau, New Mexico. *Journal of Anthropological Archaeology* 23:100-118.

Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA*, et al.* (2007) The derived FOXP2 variant of modern humans was shared with Neandertals. *Current Biology* 17(21):1908-1912.

Kretchmer N (1972) Lactose and lactase. *Scientific American* 227:71-78.

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, & Pääbo S (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19-30.

Krone SM & Neuhauser C (1997) Ancestral processes with selection. *Theoretical Population Biology* 51:210-237.

Kuhner MK, Yamato J, & Felsenstein J (1995) Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 140:1421–1430.

Kuhner MK, Yamato J, & Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the c|oalescent. *Genetics* 149:429-434.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22:768-770.

Kuhner MK (2008) Coalescent genealogy samplers: windows into population history. *Trends in Ecology and Evolution* 24(2):86-93.

Lahr MM & Foley RA (1998) Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *American Journal of Physical Anthropology* Suppl 27:137-176.

Laland KN, Odling-Smee J, & Feldman MW (1999) Evolutionary consequences of niche construction and their implications for ecology. *Proceedings of the National Academy of Sciences of the USA* 96:10242-10247.

Laland KN, Odling-Smee FJ, & Feldman MW (2001) Cultural niche construction and human evolution. *Journal of Evolutionary Biology* 14:22-33.

Laland KN & Brown GR (2002) *Sense and nonsense: evolutionary perspectives on human behaviour* (Oxford University Press, Oxford).

Laland KN, Odling-Smee FJ, & Myles S (2010) How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews Genetics* 11:137-148.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J*, et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921.

Landsteiner K (1900) Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt Bakteriologie* 27:357-362.

Laval G & Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* 20(15):2485-2487.

Lehmann L & Feldman M (2008) War and the evolution of belligerence and bravery. *Proceedings of the Royal Society B* 275(1653):2877-2885.

Leuenberger C & Wegmann D (2010) Bayesian computation and model selection without likelihoods *Genetics* 184:243–252.

Lewinsky RH, Jensen TG, Moller J, Stensballe A, Olsen J, & Troelsen JT (2005) T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Human Molecular Genetics* 14(24):3945-3953.

Lewontin RC (1972) The apportionment of human diversity. *Evolutionary Biology* 6:391-398.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, *et al.* (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319:1100-1104.

Lieberman P (2007) The evolution of human speech: its anatomical and neural bases. *Current Anthropology* 48(1):39-66.

Liljeros F, Edling CR, Amaral LAN, Stanley HE, & Åberg Y (2001) The web of human sexual contacts. *Nature* 411:907-908.

Lipo CP, Madsen ME, Dunnell RC, & Hunt T (1997) Population structure, cultural transmission, and frequency seriation. *Journal of Anthropological Archaeology* 16:301-333.

Lombard M (2008) Finding resolution for the Howiesons Poort through the microscope: micro-residue analysis of segments from Sibudu Cave, South Africa. *Journal of Archaeological Science* 35(1):26-41.

Lüning J (2005) Bandkeramische Hofplätze und absolute Chronologie der Bandkeramik. *Die Bandkeramik im 21 Jahrhundert: Symposium in der Abtei Brauweiler bei Köln*, eds Lüning J, Friedrich C, & Zimmermann A), pp 49-74.

MacEachern S, Hayes B, McEwan J, & Goddard M (2009) An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (Bos taurus) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. *BMC Genomics* 10:181.

Mace R & Pagel MD (1994) The comparative method in anthropology. *Current Anthropology* 35:549-564.

Mace R & Holden CJ (2005) A phylogenetic approach to cultural evolution. *Trends in Ecology and Evolution* 20:116-121.

Malécot G (1948) *The mathematics of heredity* (translated by Yermanos, D.M. 1969, W. H. Freeman, San Francisco).

Marean CW, Bar-Matthews M, Bernatchez J, Fisher E, Goldberg P, Herries AI*, et al.* (2007) Early human use of marine resources and pigment in South Africa during the Middle Pleistocene. *Nature* 449(7164):905-908.

Marjoram P, Molitor J, Plagnol V, & Tavaré S (2003) Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the USA* 100:15324–15328.

Marjoram P & Tavaré S (2006) Modern computational approaches for analysing molecular genetic variation data. *Nature Reviews Genetics* 7:759–770.

Maynard Smith J (1982) *Evolution and the theory of games* (Cambridge University Press, Cambridge).

Maynard Smith J (1998) *Evolutionary genetics* (Oxford University Press, Oxford).

Mayr E (1982) *The growth of biological thought* (Harvard University Press, Cambridge, MA).

McBrearty S & Brooks AS (2000) The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of Human Evolution* 39(5):453-563.

McCracken RD (1971a) Origins and implications of the distribution of adult lactase deficiency in human populations. *Journal of Tropical Pediatrics and Environmental Child Health* 17:7-10.

McCracken RD (1971b) Lactase deficiency: an example of dietary evolution. *Current Anthropology* 12:497-517.

McDougall I, Brown FH, & Fleagle JG (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433:733-736.

McElreath R & Strimling P (2008) When natural selection favors imitation of parents. *Current Anthropology* 49(2):307-316.

McElreath R, Lubell M, Richerson PJ, Waring TM, Baum W, Edsten E*, et al.* (2005) Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior* 26:483–508.

Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, Hudson RR*, et al.* (2005) Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science* 309(5741):1720-1722.

Mellars P (2004) Neanderthals and the modern human colonization of Europe. *Nature* 432:461-465.

Mellars P (2005) The impossible coincidence. A single-species model for the origins of modern human behavior in Europe. *Evolutionary Anthropology* 14(1):12-27.

Mellars P, Gravina B, & Ramsey CB (2007) Confirmation of Neanderthal/modern human interstratification at the Châtelperronian type-site. *Proceedings of the National Academy of Sciences of the USA* 104(9):3657-3662.

Menozzi P, Piazza A, & Cavalli-Sforza L (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201(4358):786-792.

Mesoudi A, Whiten A, & Laland KN (2004) Is human cultural evolution Darwinian? Evidence reviewed from the perspective of *The Origin of Species*. *Evolution* 58(1):1-11.

Mesoudi A, Whiten A, & Laland KN (2006) Towards a unified science of cultural evolution. *Behavioral and Brain Sciences* 29(4):329-347; discussion 347-383.

Mesoudi A (2008) Foresight in cultural evolution. *Biology and Philosophy* 23(243-255).

Mesoudi A & O'Brien MJ (2008) The cultural transmission of Great Basin projectile-point technology I: An experimental simulation. *American Antiquity* 73(1):3–28.

Mesoudi A & Lycett SJ (2009) Random copying, frequency-dependent copying and culture change. *Evolution and Human Behavior* 30(1):41-48.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, & Teller E (1953) Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21(6).

Miller GS, Jr. (1927) The rodents of the genus Plagiodontia. *Proceedings of the United States National Museum* 72:1-8.

Mithen SJ (1996) *The prehistory of the mind: a search for the origins of art, religion, and science* (Thames and Hudson, London).

Moran PAP (1958) Random processes in genetics. *Proceedings of the Cambridge Philosophical Society* 54:60-71.

Moran PAP (1962) *The statistical processes of evolutionary theory* (Clarendon Press, Oxford).

Morgan LH (1877) *Ancient society* (Henry Holt, New York)

Mulcare CA, Weale ME, Jones AL, Connell B, .Zeitlyn D, Tarekegn A*, et al.* (2004) The T allele of a single-nucleotide polymorphism 13.9 kb upstream of the lactase gene (LCT) (C-13.9kbT) does not predict or cause the lactase-persistence phenotype in Africans. *American Journal of Human Genetics* 74:1102-1110.

Mullis K & Faloona F (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology* 155:335-350.

Nagylaki T (1989) Gustave Malécot and the transition from classical to modern population genetics. *Genetics* 122:253-268.

Nei M & Roychoudhury AK (1993) Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution* 10(5):927-943.

Nei M & Saitou N (1986) Genetic relationship of human populations and ethnic differences in reaction to drugs and food. *Progress in Clinical Biological Research* 214:21-37.

Neiman FD (1990) An evolutionary approach to archaeological inference: aspects of architectural variation in the 17th-century Chesapeake. (Yale University, New Haven, CT).

Neiman FD (1995) Stylistic variation in evolutionary perspective: inferences from decorative diversity and interassemblage distance in Illinois woodland ceramic assemblages. *American Antiquity* 60:7–36.

Neuenschwander S, Largiadèr CR, Ray N, Currat M, Vonlathen P, & Excoffier L (2008) Colonization history of the Swiss Rhine basin by the bullhead (Cottus gobio):

inference under a Bayesian spatially explicit framework. *Molecular Ecology* 17:757-772.

Neuhauser C & Krone SM (1997) The genealogy of samples in models with selection *Genetics* 145:519-534.

Neukum G & Ivanov BA (1994) Crater size distributions and impact probabilities on Earth from lunar, terrestrial-planet, and asteroid cratering data. *Hazards due to comets and asteroids*, ed Gehrels T (University of Arizona Press, Tuscon, AZ).

Newman MEJ & Eble GJ (1999) Power spectra of extinction in the fossil record. *Proceedings of the Royal Society B* 266:1267-1270.

Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46:323-351.

Nielsen R & Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158:885-896.

Nielsen R & Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology* 18(6):1034-1047.

Nordborg M (1998) On the probability of Neanderthal ancestry. *American Journal of Human Genetics* 63:1237-1240.

Nordborg M (2001) Coalescent theory. *Handbook of statistical genetics*, eds Balding DJ, Bishop M, & Cannings C (John Wiley & Sons, Chichester), pp 179-212.

Novembre J & Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics* 40(5):646-649.

Novembre J & Di Rienzo A (2009) Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics* 10:745-755.

Nowak MA, Sasaki A, Taylor A, & Fudenberg D (2004) Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428:646-650.

Nowak MA (2006) Five rules for the evolution of cooperation. *Science* 314:1560-1563.

Nowak RM (1999) *Walker's Mammals of the World* (The John Hopkins University Press, Baltimore, MA) 6th Ed.

O'Brien MJ & Lyman RL (2000) *Applying evolutionary archaeology: a systematic approach* (Kluwer Academic, New York).

O'Brien MJ & Lyman RL (2002) Evolutionary archeology: current status and future prospects. *Evolutionary Anthropology* 11:26-36.

O'Connell JF & Allen J (2007) Pre-LGM Sahul (Pleistocene Australia-New Guinea) and the archaeology of early modern humans. *Rethinking the human revolution*, eds Mellars P, Boyle K, Bar-Yosef O, & Stringer C (McDonald Institute for Archaeological Research, University of Cambridge, Cambridge), pp 395-410.

O'Brien MJ & Lyman RL (2003) *Cladistics and Archaeology* (University of Utah Press, Salt Lake City).

O'Connell JF & Allen J (2004) Dating the colonization of Sahul (Pleistocene Australia-New Guinea): a review of recent research. *Journal of Archaeological Science* 31(6):835-853.

Odling-Smee FJ, Laland KN, & Feldman MW (2003) *Niche construction: The neglected process in evolution* (Princeton University Press, Princeton, NJ).

Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246(5428):96-98.

Ohta T & Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable molecules in a finite population. *Genetical Research* 22:201-204.

Okasha S (2006) *Evolution and the levels of selection* (Oxford University Press, Oxford).

Olds LC & Sibley E (2003) Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. *Human Molecular Genetics* 12:2333-2340.

Orlando L, Darlu P, Toussaint M, Bonjean D, Otte M, & Hänni C (2006) Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Current Biology* 16(11):R400-R402.

Ormerod P (1998) *Butterfly economics: a new general theory of social and economic behavior* (Faber and Faber, London).

Ormerod P (2005) *Why most things fail: evolution, extinction, and economics* (Faber and Faber, London).

Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N*, et al.* (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics* 38:645-679.

Panchal M & Beaumont MA (2007) The automation and evaluation of nested clade phylogeographic analysis. *Evolution* 61:1466-1480.

Pareto V (1896) *Cours d'economie politique* (Droz, Geneva, Switzerland).

Patin E, Laval G, Barreiro LB, Salas A, Semino O, Santachiara-Benerecetti S*, et al.* (2009) Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genetics* 5(4):e1000448.

Pavúk J (2005) Typologische Geschichte der Linearbandkeramik. *Die Bandkeramik im 21 Jahrhundert: Symposium in der Abtei Brauweiler bei Köln*, eds Lüning J, Friedrich C, & Zimmermann A), pp 17-39.

Peck JR (1996) Limited dispersal, deleterious mutations and the evolution of sex. *Genetics* 142(3):1053-1060.

Peck JR, Barreau G, & Heath SC (1997) Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* 145(4):1171-1199.

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R*, et al.* (2007) Diet and the evolution of human amylase gene copy number variation. *Nature Genetics* 39:1256-1260.

Petit RJ (2007) The coup de grace for the nested clade phylogeographic analysis? *Molecular Ecology* 17:516–518.

Petit RJ (2008) On the falsifiability of the nested clade phylogeographic analysis method. *Molecular Ecology* 17:1404.

Petraglia MD (2007) Mind the gap: factoring the Arabian peninsula and the Indian subcontinent into Out of Africa models. *Rethinking the human revolution*, eds Mellars P, Boyle K, Bar-Yosef O, & Stringer C (McDonald Institute for Archaeological Research, University of Cambridge, Cambridge), pp 383-394.

Petrie WMF (1899) Sequences in prehistoric remains. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 29:295-301.

Pinhasi R, Fort J, & Ammerman AJ (2005) Tracing the origin and spread of agriculture in Europe. *PLoS Biology* 3:e410.

Pinker S (1997) *How the mind works* (W. W. Norton, New York).

Pitt Rivers AH (1875) On the evolution of culture. *Journal of the Anthropological Institute* 4:293-308.

Plagnol V & Wall JD (2006) Possible ancestral structure in human populations. *PLoS Genetics* 2(7):e1050972.

Pluzhnikov A & Donnelly P (1996) Optimal sequencing strategies for surveying molecular genetic diversity. *Genetics* 144:1247-1262.

Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS*, et al.* (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443(7108):167-172.

Powell A, Shennan S, & Thomas MG (2009) Late Pleistocene demography and the appearance of modern human behavior. *Science* 324(5932):1298-1301.

Price GR (1970) Selection and covariance. *Nature* 227(5257):520-521.

Price GR (1972) Extension of covariance selection mathematics. *Annals of Human Genetics* 35:485-490.

Pritchard JK, Seielstad MT, Perez-Lezaun A, & Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* 16(12):1791–1798.

Prugnolle F, Manica A, & Balloux F (2005) Geography predicts neutral genetic diversity of human populations. *Current Biology* 15(5):R159-R160.

Pybus OG, Rambaut A, & Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155:1429-1437.

R Development Core Team (2009) R: A language and environment for statistical computing. *R Foundation for Statistical Computing* http://www.R-project.org.

Ramachandran S, et al., (2005) Support from the relationship between geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the USA* 102:15942-15947.

Ramakrishnan U & Hadly EA (2009) Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Molecular Ecology* 18:1310-1330.

Ray N, Currat M, & Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution* 20(1):76-86.

Ray N, Currat M, Berthier P, & Excoffier L (2005) Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Research* 15(8):1161-1167.

Ray N, Wegmann D, Fagundes NJR, Wang S, Ruiz-Linares A, & Excoffier L (2010) A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Molecular Biology and Evolution* 27(2):337–345.

Read D (2006) Tasmanian knowledge and skill: maladaptive imitation or adequate technology? *American Antiquity* 71(1).

Redner S (1998) How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B* 4:131-134.

Rendell L, Boyd R, Cownden D, Enquist M, Eriksson K, Feldman MW *et al.* (2010) Why copy others? Insights from the Social Learning Strategies Tournament. *Science* 328:208-213

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, *et al.* (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics* 59:185-203.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, *et al.* (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *American Journal of Human Genetics* 67:1251-1276.

Richerson PJ & Boyd R (1978) A dual inheritance model of the human evolutionary process I: basic postulates and a simple model. *Journal of Social and Biological Structures* 1:127-154.

Richerson PJ, Boyd R, & Bettinger RL (2001) Was agriculture impossible during the Pleistocene but mandatory during the Holocene? A climate change hypothesis. *American Antiquity* 66:387-412.

Richerson PJ & Boyd R (2005) *Not by genes alone: how culture transformed human evolution* (University of Chicago Press, Chicago).

Richerson PJ, Boyd R, & Henrich J (2010) Gene-culture coevolution in the age of genomics. *Proceedings of the National Academy of Sciences USA* 107 (suppl.2):8985-8992.

Ripley BD (1987) *Stochastic simulation* (Wiley).

Robinson WR (1951) A method for chronologically ordering archaeological deposits. *American Antiquity* 16:193-301.

Rodrigo AG & Felsenstein J (1999) Coalescent approaches to HIV population genetics. *The evolution of HIV*, ed Crandall KA (Johns Hopkins University Press, Baltimore), pp 233–272.

Rogers AR (1988) Does biology constrain culture? *American Anthropologist* 90:819-831

Rogers AR (1995) Genetic evidence for a Pleistocene population explosion. *Evolution* 49:608-615.

Rogers EM (1995) *Diffusion of innovations* (Free Press, New York, NY).

Röhrs M & Herre W (1961) Zur Frühentwicklung der Haustiere. Die Tierreste der Neolithischen Siedlung Fikirtepe am Kleinasiatischen Gestade des Bosporus. *Zeitschrift für Tierzüchtung und Züchtungsbiologie*), pp 110-127.

Rose S, Lewontin RC, & Kamin LJ (1984) *Not in our genes: biology, ideology and human nature* (Pantheon, London).

Rosen KT & Resnick M (1980) The size distribution of cities: an examination of the Pareto law and primacy. *Journal of Urban Economics* 8:165-186.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA*, et al.* (2002) Genetic structure of human populations. *Science* 298(5602):2381-2385.

Russell B (1988) Knowledge by acquaintance and knowledge by description. *Propositions and Attitudes*, eds Salmon N & Soames S (Oxford University Press Oxford).

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O*, et al.* (2006) Positive natural selection in the human lineage. *Science* 312:1614-1620.

Saiki R, Scharf S, Faloona F, Mullis K, Horn G, & Erlich H (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350-1354.

Scarre C, ed. (2005) *The human past: world prehistory and the development of human societies* (Thames & Hudson, London).

Schauer P (2008) Cultural evolution in the age of Athens: drift and selection in Greek figure-painted pottery. (University College London, London).

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE*, et al.* (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290:1155-1159.

Shapiro B, Sibthorpe D, Rambaut A, Austin J, Wragg GM, Bininda-Emonds ORP*, et al.* (2002) Flight of the Dodo. *Science* 295(1):1683.

Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV*, et al.* (2004) Rise and fall of the Beringian steppe bison. *Science* 306:1561-1565.

Shennan S & Bentley RA (2007) Style, interaction, and demography among the earliest farmers of central Europe. *Cultural Transmission in Archaeology*, ed O'Brien MJ (Society for American Archaeology, Washington DC).

Shennan SJ & Steele J (1999) Cultural learning in hominids: a behavioral ecological approach. *Mammalian Social Learning: Comparative and Ecological Perspectives*, eds Box H & Gibson K (Cambridge University Press, Cambridge), pp 367-388.

Shennan SJ (2001) Demography and cultural innovation: a model and its implications for the emergence of modern human culture. *Cambridge Archaeological Journal* 11(1):5-16.

Shennan SJ & Wilkinson JR (2001) Ceramic style change and neutral evolution: a case study from Neolithic Europe. *American Antiquity* 66:577-594.

Shennan SJ (2002) *Genes, memes and human history* (Thames & Hudson, London).

Shennan SJ (2008) Evolution in archaeology. *Annual Review of Anthropology* 37:75-91.

Sherry ST, Rogers AR, Harpending HC, Soodyall H, Jenkins T, & Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. *Human Biology* 66:761-775.

Sherry ST, Harpending HC, Batzer MA, & Stoneking M (1997) Alu evolution in human populations: using the coalescent to estimate effective population size. *Genetics* 147(4):1977-1982

Simkin MV & Roychowdhury VP (2006) An introduction to the theory of citing. *Significance* 3(4):179-181.

Simon HA (1955) On a class of skew distribution functions. *Biometrika* 42(3/4):425-440.

Simoons FJ (1970) Primary adult lactose intolerance and the milking habit: a problem in biological and cultural interrelations. II. A culture historical hypothesis. *American Journal of Digestive Disease* 15:695-710.

Simoons FJ (1978) The geographic hypothesis and lactose malabsorption. A weighing of the evidence. *American Journal of Digestive Disease* 23:963-980.

Simoons FJ (2001) Persistence of lactase activity among northern europeans: A weighing of the evidence for the calcium absorption hypothesis. *Ecology of Food and Nutrition* 40:397-469.

Sisson SA, Fan Y, & Tanaka MM (2007) Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the USA* 104(6):1760-1765.

Sjödin P, Kaj I, Krone SM, Lascoux M, & Nordborg M (2005) On the meaning and existence of an effective population size. *Genetics* 169:1061-1070.

Skyrms B (1996) *Evolution of the social contract* (Cambridge University Press, Cambridge).

Slatkin M (1994) An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research* 64:71-74.

Slatkin M (1996) A correction to the exact test based on the Ewens sampling distribution. *Genetical Research* 68:259-260.

Smith EA & Winterhalder B (1992) Natural selection and decision making: some fundamental principles. *Evolutionary Ecology and Human Behavior*, eds Smith EA & Winterhalder B (Aldine de Gruyter, New York), pp 25-60.

Smith EA (2000) Three styles in the evolutionary analysis of human behavior. *Adaptation and human behavior: an anthropological perspective*, eds Cronk L, Chagnon N, & Irons W (Aldine de Gruyter, New York), pp 27-46.

Sokal RR, Oden NL, & Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351(6322):143-145.

Soltis J, Boyd R, & Richerson PJ (1995) Can group-functional behaviors evolve by cultural group selection? *Current Anthropology* 36(3):473–494.

Soressi M & d'Errico F (2007) Pigments, gravures, parures: les comportements symboliques controversés des Néandertaliens. *Les Néandertaliens. Biologie et cultures. Document préhistoriques 23.* (Éditions du CTHS, Paris), pp 297-309.

Sornette D (2006) *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools* (Springer, Heidelberg) 2nd ed.

Sousa VC, Fritz M, Beaumont MA, & Chikhi L (2009) Approximate Bayesian computation without summary statistics: the case of admixture. *Genetics* 181:1507-1519.

Sperber D (1996) *Explaining culture: a naturalistic approach* (Blackwell, Oxford).

Sperber D & Claidiére N (2008) Defining and explaining culture (comments on Richerson and Boyd, *Not by genes alone*). *Biology and Philosophy* 23(283-292).

Steele J, Glatz C, & Kandler A (2010) Ceramic diversity, random copying, and tests for selectivity in ceramic production. *Journal of Archaeological Science* 37(6):1348-1358

Stephens M & Donnelly PJ (2000) Inference in molecular population genetics. *Journal of the Royal Statistical Society Series B* 62:1-31.

Stiner MC, Munro ND, & Surovell TA (2000) The tortoise and the hare: small-game use, the broad-spectrum revolution, and Paleolithic demography. *Current Anthropology* 41(1):39-73.

Stiner MC & Kuhn SL (2006) Changes in the 'connectedness' and resilience of Paleolithic societies in Mediterranean ecosystems. *Human Ecology* 34:693–712.

Strabo (1969) *The Geography of Strabo in eight volumes* (English trans. by H. L. Jones).

Strien H-C (2000) *Untersuchungen zur Bandkeramik in Württemberg* (Habelt, Bonn).

Strimling P, Sjöstrand J, Enquist M, & Eriksson K (2009) Accumulation of independent cultural traits. *Theoretical Population Biology* 76(2):77-83.

Strimmer K & Pybus OG (2001) Exploring the demographic history of DNA sequences using the generalized skyline plot. *Molecular Biology and Evolution* 18(12):2298-2305.

Stringer C & Andrews P (1988) Genetic and fossil evidence for the origin of modern humans. *Science* 239(4845):1263-1268.

Stringer C (2000) Coasting out of Africa. *Nature* 405:24-25.

Stringer C (2002) Modern human origins: progress and prospects. *Philosophical Transactions of the Royal Society B* 357:563-579.

Stringer C (2007) Biological and demographic perspectives on modern human origins. *Rethinking the human revolution*, eds Mellars P, Boyle K, Bar-Yosef O, & Stringer C (McDonald Institute for Archaeological Research, University of Cambridge, Cambridge).

Sun C, Kong QP, Palanichamy MG, Agrawal S, Bandelt HJ, Yao YG*, et al.* (2006) The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Molecular Biology and Evolution* 23(3):683-690.

Swallow DM (2003) Genetics of lactase persistence and lactose intolerance. *Annual Review of Genetics* 37:197-219.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437-460.

Tavaré S (1984) Lines of descent and genealogical processes, and their application in population genetics models. *Theoretical Population Biology* 26:119-164.

Tavaré S, Balding DJ, Griffiths RC, & Donnelly P (1997) Inferring coalescence times from DNA sequence data. *Genetics* 145:505–518.

Tehrani JJ & Collard M (2002) Investigating cultural evolution through biological phylogenetic analyses of Turkmen textiles. *Journal of Anthropological Archaeology* 21(4):443-463.

Tehrani JJ & Collard M (2009) On the relationship between interindividual cultural transmission and population-level cultural diversity: a case study of weaving in Iranian tribal populations. *Evolution and Human Behavior* 30:286–300.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* 7(4):381-397.

Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology* 13:789-809.

Templeton AR (2008) Nested clade analysis: an extensively validated method for strong phylogeographic inference. *Molecular Ecology* 17:1877-1880.

Templeton AR (2009a) Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographical analysis vs. approximate Bayesian computation. *Molecular Ecology* 18:319-331.

Templeton AR (2009b) Why does a method that fails continue to be used? The answer. *Evolution* 63:807-812.

Templeton AR (2010) Coalescent-based, maximum likelihood inference in phylogeography. *Molecular Ecology* 19(3):431-435.

Texier J-P, Porraz G, Parkington J, Rigaud J-P, Poggenpoel C, Miller C*, et al.* (2010) A Howiesons Poort tradition of engraving ostrich eggshell containers dated to 60,000 years ago at Diepkloof Rock Shelter, South Africa. *Proceedings of the National Academy of Sciences USA* 107(14):6180-6185.

The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449:851-862.

Thomas MG, Stumpf MPH, & Harke H (2006) Evidence for an apartheid-like social structure in early Anglo-Saxon England. *Proceedings of the Royal Society B* 273:2651-2657.

Thorne A & Wolpoff MH (1981) Regional continuity in Australasian Pleistocene hominid evolution. *American Journal of Physical Anthropology* 55:337-349.

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Cheung K, Kidd JR*, et al.* (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 271(5254):380-1387.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS*, et al.* (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* 39:31-40.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A*, et al.* (2009) The genetic structure and history of Africans and African Americans. *Science* 324:1035-1044.

Tomasello M (1999) *The cultural origins of human cognition* (Harvard University Press, Cambridge, MA).

Toni T, Welch D, Strelkowa N, Ipsen A, & Stumpf MP (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface* 6:187-202.

Toni T & Stumpf MPH (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26(1):104–110.

Tremblay M & Vezina H (2000) New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *American Journal of Human Genetics* 66:651-658.

Tresset A (1996) Le rôle des relations homme/animal dans l'évolution économique et culturelle des sociétés des V-VI millénaires en Bassin Parisien. (Université de Paris I, Panthéon-Sorbonne, Paris).

Tresset A (1997) L'approvisionnement carné Cerny dans le contexte néolithique du Bassin Parisien. *La Culture de Cerny: nouvelle économie, nouvelle société au Néolithique*, eds Constantin C, Mordant D, & Simonin DActes du Colloque International de Nemours), pp 299-314.

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P*, et al.* (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410:1088-1091.

Turvey ST (2009) *Holocene extinctions* (Oxford University Press, Oxford).

Tylor EB (1871) *Primitive culture* (John Murray, London)

Vanhaeren M (2005) Speaking with beads: the evolutionary significance of personal ornaments. . *From Tools to Symbols from Early Hominids to Humans*, eds d'Errico F & Blackwell L (Wits University Press, Johannesburg), pp 525-553.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG*, et al.* (2001) The sequence of the human genome. *Science* 291:1304-1351.

Vigilant L, Stoneking M, Harpending HC, Hawkes K, & Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253(5027):1503-1507.

Vigne J-D (2006) Maîtrise et usages de l'élevage et des animaux domestiques au Néolithique: quelques illustrations au Proche-Orient et en Europe. *Populations néolithiques et environnements*, ed Guilaine J (Errance éd., Paris), pp 87-114.

Vigne J-D & Helmer D (2007) Was milk a "secondary product" in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. *Anthropozoologica* 42:9-40.

Vigne J-D (2008) Zooarchaeological aspects of the Neolithic diet transition in the Near East and Europe, and their putative relationships with the Neolithic Demographic Transition. *The Neolithic Demographic Transition and its consequences*, eds Bocquet-Appel J-P & Bar-Yosef O (Springer Verlag, New York), pp 179-205.

Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, & Rienzo AD (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proceedings of the National Academy of Sciences of the USA* 102(51):18508-18513.

Voight BF, Kudaravalli S, Wen X, & Pritchard JK (2006) A map of recent positive selection in the human genome *PLoS Biology* 4(3):e72.

Wakeley J (2009) *Coalescent theory: an introduction* (Roberts and Co. Publishers, Greenwood Village, CO).

Wand M & Ripley B (2009) KernSmooth: Functions for kernel smoothing for Wand & Jones (1995). http://CRAN.R-project.org/package=KernSmooth.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7:256-275.

Watterson GA (1977) Heterosis or neutrality? *Genetics* 85:789-814.

Watterson GA (1978) The homozygosity test of neutrality. *Genetics* 88:405-417.

Wegmann D, Leuenberger C, & Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*.

Weiss G & von Haeseler A (1998) Inference of population history using a likelihood approach. *Genetics* 149:1539-1546.

Weiss KM (2004) The unkindest cup. *Lancet* 363:1489-1490.

White TD, Asfaw B, DeGusta D, Gilbert H, Richards GD, Suwa G, *et al.* (2003) Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* 423(6941):742-747.

Willerslev E & Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society B* 272:3-16.

Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, & Nielsen R (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genetics* 3(6):e90.

Willis JC & Yule GU (1922) Some statistics of evolution and geographical distribution in plants and animals, and their significance. *Nature* 109:177-179.

Wilson IJ & Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150:499-510.

Wilson EO (1975) *Sociobiology: the new synthesis* (Harvard University Press, Cambridge MA).

Woods CA & Howland EB (1979) Adaptive radiation of capromyid rodents: anatomy of the masticatory apparatus. *Journal of Mammalogy* 60:95-116.

Wolpoff MH, Hawks J, & Caspari R (2000) Multiregional, not multiple origins. *American Journal of Physical Anthropology* 112:129-136.

Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97-159.

Wright S (1943) Isolation by distance. *Genetics* 28:114-138.

Yule GU (1925) A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London B* 213:21-87.

Zilhão J & d'Errico F (1999) The chronology and taphonomy of the earliest Aurignacian and its implications for the understanding of Neandertal extinction. *Journal of World Prehistory* 13:1-68.

Zilhão J, d'Errico F, Bordes J-G, Lenoble A, Texier J-P, & Rigaud J-P (2006) Analysis of Aurignacian interstratification at the Châtelperronian-type site and implications for the behavioral modernity of Neandertals. *Proceedings of the National Academy of Sciences of the USA* 103(33):12643-12648.

Zilhão J (2007) The emergence of ornaments and art: an archaeological perspective on the origins of "behavioral modernity". *Journal of Archaeological Research* 15:1-54.

Zipf GK (1929) Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 15:1-95.

Zipf GK (1949) *Human behaviour and the principle of least effort* (Addison-Wesley, Reading, MA).

Zvelebil M & Zvelebil KV (1988) Agricultural transition and Indo-European dispersals. *Antiquity* 62:574-583.

# Appendix

**Table A1:** Southwest German LBK data phase durations (see Strien 2000, Shennan and Bentley (2007).

| phase | 1 | 2A | 2B/3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| estimated duration (house generations) | ~5 | ~3 | ~5 | 2 | 1 | 2 | 1 | 1 |

**Table A2:** Southwest German LBK decorative motif frequency data (see Strien 2000, Shennan and Bentley 2007).

| variant ID | max. freq. | gens to max. freq. | years to max. freq. |
|---|---|---|---|
| 1 | 0.013513514 | 1 | 25 |
| 2 | 0.094594595 | 2.5 | 62.5 |
| 3 | 0.021126761 | 2 | 50 |
| 4 | 0.006756757 | 2.5 | 62.5 |
| 5 | 0.047297297 | 2.5 | 62.5 |
| 6 | 0.006756757 | 2.5 | 62.5 |
| 7 | 0.006756757 | 2.5 | 62.5 |
| 8 | 0.006756757 | 2.5 | 62.5 |
| 9 | 0.297297297 | 3.5 | 87.5 |
| 10 | 0.007042254 | 2 | 50 |
| 11 | 0.067567568 | 3.5 | 87.5 |
| 12 | 0.141891892 | 4.75 | 118.75 |
| 13 | 0.027027027 | 3.5 | 87.5 |
| 14 | 0.040540541 | 4.75 | 118.75 |
| 15 | 0.007042254 | 2 | 50 |

**Table A2** (continued)

| variant ID | max. freq. | gens to max. freq. | years to max. freq. |
|---|---|---|---|
| 16 | 0.154929577 | 4.25 | 106.25 |
| 17 | 0.176056338 | 4.25 | 106.25 |
| 18 | 0.056338028 | 4.25 | 106.25 |
| 19 | 0.024159664 | 2.5 | 62.5 |
| 20 | 0.013513514 | 3.5 | 87.5 |
| 21 | 0.007042254 | 2 | 50 |
| 22 | 0.091549296 | 5.5 | 137.5 |
| 23 | 0.030462185 | 3.75 | 93.75 |
| 24 | 0.004201681 | 2.5 | 62.5 |
| 25 | 0.082983193 | 3.75 | 93.75 |
| 26 | 0.174657534 | 2.75 | 68.75 |
| 27 | 0.020408163 | 2.25 | 56.25 |
| 28 | 0.267123288 | 4 | 100 |
| 29 | 0.050632911 | 6.5 | 162.5 |
| 30 | 0.050632911 | 6.5 | 162.5 |
| 31 | 0.240506329 | 6.5 | 162.5 |
| 32 | 0.164556962 | 6.5 | 162.5 |
| 33 | 0.208163265 | 9.25 | 231.25 |
| 34 | 0.141843972 | 5.25 | 131.25 |
| 35 | 0.139240506 | 10.5 | 262.5 |
| 36 | 0.028368794 | 8 | 200 |
| 37 | 0.028368794 | 8 | 200 |
| 38 | 0.546099291 | 8 | 200 |
| 39 | 0.028368794 | 8 | 200 |
| 40 | 0.184397163 | 8 | 200 |
| 41 | 0.021276596 | 8 | 200 |
| 42 | 0.021276596 | 8 | 200 |